

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:07:17 ; Search time 15810.8 Seconds
(without alignments)
2705.363 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcgaggtccactgtgct.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb.ba.*

2: gb.btg.*

3: gb.in.*

4: gb.om.*

5: gb.ov.*

6: gb.pat.*

7: gb.ph.*

8: gb.pl.*

9: gb.pr.*

10: gb.ro.*

11: gb.sts.*

12: gb.sy.*

13: gb.un.*

14: gb.vi.*

15: em.ba.*

16: em.fun.*

17: em.hum.*

18: em.in.*

19: em.mu.*

20: em.cm.*

21: em.or.*

22: em.ov.*

23: em.pat.*

24: em.ph.*

25: em.pl.*

26: em.ro.*

27: em.sts.*

28: em.un.*

29: em.vi.*

30: em.htg.hum.*

31: em.htg.inv.*

32: em.htg.other.*

33: em.htgo.inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description

ALIGNMENTS

RESULT	1	AF176838	2170 bp	mrna	linear	PRI 22-SEP-1999
LOCUS	AF176838	Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-St) mRNA, complete cds.				
DEFINITION	AF176838					
ACCESSION	AF176838					
VERSION	AF176838.1	GI:5917705				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	1 (bases 1 to 2170)					
AUTHORS	Lee,J.K., Bhakta,S., Rosen,S.D. and Hemmerich,S.					
TITLE	N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue					
JOURNAL	Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)					
MEDLINE	99423499					
REFERENCE	2 (bases 1 to 2170)					
AUTHORS	Lee,J.-K., Bhakta,S., Rosen,S.D. and Hemmerich,S.					
TITLE	Direct Submission					
JOURNAL	Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,					

1	1973	96.5	2170	9	AF176838	Homo sapi
2	1774.6	85.8	3278	9	AF246718	Homo sapi
3	1753.6	85.8	157337	9	AC009163	Homo sapi
4	1753.6	85.8	194832	9	AC025287	Homo sapi
5	1742.6	85.3	159072	2	AC026419	Homo sapi
6	1719	84.1	71503	9	AF219991	Homo sapi
7	1248.6	61.1	1462	9	AF176839	Homo sapi
8	1101	53.9	3786	9	AF280086	Homo sapi
9	1052	51.5	2544	9	AF219990	Homo sapi
10	1046.4	51.2	159072	2	AC026419	Homo sapi
11	1046.4	51.2	208185	2	AC009105	Homo sapi
12	947	46.3	1647	6	AX327330	Sequence
13	751.2	36.8	1989	10	AF176840	Mus muscu
14	750.4	36.7	1740	10	AF176841	Mus muscu
15	719.6	35.2	139252	2	AC095664	Rattus no
16	571.2	27.9	90256	2	AP001582	AP001582 Homo sapi
17	568	27.8	215647	2	AC068591	AC068591 Homo sapi
18	556.8	27.2	174187	2	AP002792	AP002792 Homo sapi
19	542.6	26.5	137499	2	AC015931	AC015931 Homo sapi
20	400	19.6	2032	9	AF131235	AF131235 Homo sapi
21	398.4	19.5	1333	9	AF149783	AF149783 Homo sapi
22	398.4	19.5	1992	9	AF280088	AF280088 Homo sapi
23	398.4	19.5	2011	9	AK026635	AK026635 Homo sapi
24	398.4	19.5	183228	9	AC010547	AC010547 Homo sapi
25	358.8	17.6	2201	10	AF109155	AF109155 Mus muscu
26	354	17.3	1926	10	AF131236	AF131236 Mus muscu
27	235.2	11.5	2156	6	AR071396	AR071396 Sequence
28	235.2	11.5	2156	6	EA4937	EA4937 Human mRNA
29	235.2	11.5	2731	9	AB012192	AB012192 Homo sapi
30	235.2	11.5	6961	9	AB017915	AB017915 Homo sapi
31	235.2	11.5	178251	9	AC073370	AC073370 Homo sapi
32	235.2	11.5	196465	9	AC022392	AC022392 Homo sapi
33	228.8	11.2	98392	2	AL590620	AL590620 Homo sapi
34	228.8	11.2	172837	9	AL450304	AL450304 Human DNA
35	226.2	11.1	103567	9	HS12406	AL021326 Homo sapi
36	224.4	11.0	54666	9	AC073487	AC073487 Homo sapi
37	223	10.9	176895	2	AC097065	AC097065 Homo sapi
38	222	10.9	140210	2	AC002993	AC002993 Homo sapi
39	221.6	10.8	183974	9	AC093421	AC093421 Homo sapi
40	220.8	10.8	151068	9	AL133405	AL133405 Human DNA
41	220.6	10.8	148463	2	AC012140	AC012140 Homo sapi
42	220.6	10.8	207408	2	AC092872	AC092872 Pan trogl
43	220.6	10.8	209476	2	AC021027	AC021027 Homo sapi
44	220.4	10.8	57573	2	AC087509	AC087509 Homo sapi
45	220.4	10.8	180317	2	AC025590	AC025590 Homo sapi

FEATURES	source	3401 Hillview Avenue, Palo Alto, CA 94304, USA
Location/Qualifiers	1. .2170	
/organism="Homo sapiens"	/db.xref="taxon.9606"	
/chromosome="16"	/map="16q23.1"	
/tissue_type="intestine"	1. .2170	
/gene	/gene="I-GlcNAC-6-Sn"	
CDS	344. 1516	
/note="possibly involved in biosynthesis of mucosal ligands for L-selectin"	/codon_start=1	
/product="N-acetylglycosamine 6-O-sulfotransferase"	/protein_id="AAD56000.1"	
/db.xref="GI-5917706"	/translation="MWLPKPSSTVTVILLAQTCLLLTIIIRPSPSSPAGGSDRVHV LVLSNRSGSSFGQLFSDPQVFLMEPAWHVWTTLSQCSAATLMAVRDLMRSLF CDMDVEDATYMPQGRNLSAFENFWATISRALGSPFASFPFGTI SKQDVCTLCITRQFPI LAREACRSYSHVLKWEVFNQLVPLSDPALNRIVHLVRDPAVLRSRAAGPI I LARDNGILVGTNGKWEADP HLIRIEVGRSHVRIEAAATLKPPPTLRGYRLVRFD LAREPLAEIPALYAFPTGLITPOLEARNIHNHSGSIGKPIEAFHSSSNARNVDSQW RHALPTEKILRVOECAGALQILGRIPIVSADQQRDLILDLVLRPGDFHFSWASPD"	
BASE COUNT	449 a 682 c 612 g 427 t	
ORIGIN		
Query Match	96.5%;	Score 1973; DB 9; Length 2170;
Best Local Similarity	99.7%;	Pred. No. 6.3e-403;
Matches 1998:	Conservative 0; Mismatches 5; Indels 2; Gaps 3;	
QY	41	gaaagctccagtgaggaactggtctctgagaaactctgtggacatgagtgattcaac 100
Db	167	GAAAGCTCCCAAGTGAAGAACTGGCTTCCTGGAGACTCTGTGGCATAGAGTGATTCAC 225
QY	101	caacttaagaagaacctbtggtttcccttggaacacagatgtcgagacatctcccattgatt 160
Db	227	CACCTTAAGAAGAACTCTGGCTTTCTCTGGAAACACAGATGTCGAGACATCTCCCATGGATT 286
QY	161	tgtgatcagcgttgacgtctccacagacacctggaagctggccccccagcgcgcgcgatg 220
Db	287	TGTGATCAGCGTTGCACGCTTCCACAGACGCCCTTGACGGTGGCCCCCAGCGCCGCCGATG 346
QY	221	tggctgcacgggttctccagcaagacagtgacagtgctctcctgtggacacagacacctgc 280
Db	347	TGGCTGCACGGTTCTCCACGACAGACAGTGCACAGTGTCTCTCTCTGGCAGACAGACCTGC 406
QY	281	ctcgtgctcttcacatctcccggccaggcgccctcatccccagccggcgcgaggatcgt 340
Db	407	CTCGTGTCTTTCATCATCTCCGSCCAGGGCCCTCATCCCCAGCGCGCGAGGATCGT 466
QY	341	gtcacagctgctgtgtctcctgctggtcgtgggctcattcctcttgggcagcgtcttc 400
Db	467	GTCACAGTGTGTGGTGTGTCTCTGTGGCGTCCGGGCTCATCCCTCTTGTGGCCGAGCTTTC 526
QY	401	agccagcaccocagcgtctcttacctgatggagccgcgtggcatgtgtgagacacacctg 460
Db	527	AGCCAGCACCCGACGTCCTTCTACCTGATGGAGCCCGCTGGGCATGTGTGGACACACCTG 586
QY	461	tgcgagggcagcgggcaaacgctgcacatggcctgctgcgacacctgatgcgtctctatcttt 520
Db	587	TCGACAGGCACAGCGGGCAACGCTGCACATGGCGTGGCGACCTGATGGCTCTATCTTT 646
QY	521	tttgtgacatgacggtgtttgatgcctacatgccacagagccgaacctgttcgcctttt 580
Db	647	TTTGTGCACATGACGACGTGTGTTGATGGCTTACATGCCACAGAGCCGAACACCTGTCCGCTTT 706
QY	581	tccaactgggcaacgagccgcgcgtgtgtctgcgcgcgcgcctgtcagcgcttttcccca 640
Db	707	TTCAACTTGGCAGAGCGCGCGCTGTGTCTCGCGCGCCGCCCTTCACAGCGCTTTTCCCGGA 766

QY	1037	ttcagagacctggcgcgaggacgcgtggcagagatccgcgacactacgccttcaccggc	1039
Db	2279	TTCCAGGACCTTGGCGCGGAGCCGCTGGCGAGAGATCCGCGGCACCTCTACGCCCTTACCGGCG	2338
QY	1097	ctgaccttcacgcacagcttcgaggccttgatcccaacatcacccacgggtcggggtc	1156
Db	2339	CTGACCTTCACGCCACAGCTTCGAGGCGCTGGATCCCAACATCACCCACGGGTGCGGGATC	2398
QY	1157	ggcaagccaatcagcgcttcacactcttgtctagaaatggcgcaaatgtccacagcc	1216
Db	2399	GGCAAGCCAAATCAGGCGCTTCCATATCTCGTTAGGAAATGGCGCGACAGCTTCCACAGGCC	2458
QY	1217	tggcgccacgcttgcccttcactaaagatccctgcgcgtgcaggaggtgtgcccgcgcgcg	1276
Db	2459	TGGCGCCACGCGTTGCGCCTTCACTAAGATCCTGCGCGTGCAGGAGGTGTGCGCGCGCGCG	2518
QY	1277	ctgcagctgctgggctaccggcctgtgtactctgcgcacacagcagcgtgaacctcacctg	1336
Db	2519	CTGCAGCTGCTGGGCTACCGGCGCTGGTACTCTGCGGACCCAGCAGCGTGTACCTCACCGCTG	2578
QY	1337	gatctgtctgcacagcgagccacaccacttcagctgggcattgcctgcctgactgagaactc	1396
Db	2579	GATCTGCTGCTGCCACGAGGCCACAGCCACTTTCAGCTGGGCGATCCGCTGACTGAGAACTC	2638
QY	1397	tgggccttaagacaagcccgcaactgtggtgcgcagggcccgaggagcagatcgaatgtgg	1456
Db	2639	TGGGCGCTTAGACAGCGCCCGCAACTGTGTGCCAGGCCAGGAGGCGACTGCATGCTGG	2698
QY	1457	aaaagagctggggccatggggaa-caagtcctactatcaaccggaggtttggggctc	1515
Db	2699	AGAGGGAGCTTGGGCGCATTGGGAGACGAGTCCCTTACTATCAACCGGGAGTTTGGGGTCC	2758
QY	1516	tcacctgaagtgaagcaaggactgcagtttctctctcctgattctcgggttttcctttg	1575
Db	2759	TCGCTCTCAAGTAGGCAAGGACTGCAGTTCCTTCTCTCCGTATCTCGGTTTCCTTTG	2818
QY	1576	agttctctggagctgccttctcatcaggtgcactcttcattggaagaaacactcttgccc	1635
Db	2819	AGTCTCTGGAGCTGCGCTTCTCATCAGGTGCACCTTTCATGG-AAAGCAACTCTTGCCCC	2877
QY	1636	tacctctctggcgagggagtaagtactgctaaattaaattaaatgtgtccagggcc	1695
Db	2878	TACCTCTCTGGGCGCAGGAGTAGTTCCTGCTAAATTAATTAATGTGTGCCAGGCC	2937
QY	1696	gggtgcggtggtctatgcctgtaatccacagattttgagaggtgcgcggttggaatcac	1755
Db	2938	GGGTGCGGTGGCTCATGCTGTAAATCCAGCATTTTGAGAGGCTTGAGCGGGTGGATCAC	2997
QY	1756	ctgagctcagga-ttcaaaacacgctggcacaatagtgaaacccctctctactaaa	1814
Db	2998	CTGAGTCTAGGAGTTCGAAACACAGCTGGCCACATAGTGAACCCCGCTCTCTACTATAAA	3057
QY	1815	atgcaaaaattagtcgcgcgtgtgtgcacactctctgtaatccagctacttaggagctg	1874
Db	3058	ATGCAAAAATTAGTCCGCGTGGTGGCACACTCTGTATNTCCAGCTACTTAGAGGCTG	3117
QY	1875	aggtgggaaaaatcaattggactccaaagttggaggttgcagtaagctgaaatccatgccac	1934
Db	3118	AGGTGGGAAATCACTTGGACTCCAGAGGTGGAGTTCAGTAAGCTGAGATCATGCCAC	3177
QY	1935	tgcacctagcttgggtggcaaaccaaacctctatcaaaaaataataataattggt	1994
Db	3178	TGCACCTTAGCTTGGTGGCGACGACAGCTCTATCAAAAAATAAGTAATAAATTTGTT	3237
QY	1995	caaaagtctgcgcaaaaaaataaaaaaataaaaaaataaaaaa	2035
Db	3238	CAAAAGTCGAAAAAATAAAAAAATAAAAAAATAAAAAA	3278
RESULT	3		
AC009163/c			
LOCUS	AC009163	157337 bp	DNA linear

RESULT 3
AC009163/c
LOCUS

DEFINITION	Homo sapiens chromosome 16 clone RP11-77K12, complete sequence.
ACCESSION	AC009163
VERSION	AC009163.5 GI:16596526
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	DOE Joint Genome Institute.
TITLE	1 (bases 1 to 157337)
JOURNAL	DOE Joint Genome Institute.
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 157337)
TITLE	DOE Joint Genome Institute.
JOURNAL	Direct Submission
REFERENCE	Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS	3 (bases 1 to 157337)
TITLE	DOE Joint Genome Institute.
JOURNAL	Direct Submission
REFERENCE	Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS	On Nov 2, 2001 this sequence version replaced gi:7689976.
TITLE	Sequence Quality Assessment:
JOURNAL	This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
COMMENT	All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
SOURCE

```
FEATURES
source
Location/Qualifiers
1.157337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-77K12"
```

BASE COUNT	45295 a	36632 c	36308 g	39102 t
ORIGIN	/clone="RPII-77K12"			

Query Match	85.8%;	Score 1753.6;	DB 9;	Length 157337;
Best Local Similarity	99.1%;	Pred. No. 3.2e-268;		
Matches 1795;	Conservative 0;	Mismatches 14;	Indels 3;	Gaps 3;
QY 197	cggTggccccagccgcccacatgctgcacacggttctccagcagacagtcacagtg	256		
Db 144606	CGRAGGCCCCACGCGCCCGCATGTGGCTGGACGGTTCTCCAGACACAGTGCACAGTG	144547		
QY 257	ctctctctggcacagacacactgctctgctcttcacatctctccggccagggccctcca	316		
Db 144546	CTCCTCTCTGGCACAGACACCTGCCTCTGTCTTCATCATCTCCCGGCCAGGCCCTCA	144487		
QY 317	tcccccagccggccgacagatcgtgtacacgctgctgtgtcctcgtgcgcctcgagcc	376		
Db 144486	TCCCCAGCCGCGCGCAGAGATCGTGTGTGTGTGTGTCTCTGTGGCGCTCGGGC	144427		
QY 377	tcaactcttctggggccagctcttccagccagacaccccgacgtctcttaccctgatggagccc	436		
Db 144426	TCACTCTCTTGTGGCCAGCAGCTTCACGACGACACCCGACGCTCTCACTCATGCGAGGCC	144367		

QY 437 gcttgatgtgtgacacactgttcagggcagcggcggaacactgcaaatggccctg 496
|||||
Db 144366 GGGTGGCATGTGTGACCAACCTGTGCGAGGCGAGCGCGCAACGCTGCACATGGCCGTG 144307
|||||
QY 497 cgcgacctgactgcctctatcttcttggcgacacagacgtgtttgatccctacatgccca 556
|||||
Db 144306 CCGGACCTGATCGGCTCTATCTTTTGTGCGACATGGACGCTGTTTGTATGCCATACATGCCA 144247
|||||
QY 557 cagagccgaacactgtccgccttttcttaactgggaacagagccgcgtgtgtgtgcgcg 616
|||||
Db 144246 CAGAGCGGAACCTGTCCGCCCTTTTCAACTGGCAACAGAGCGCGCGTGTCTCGCG 144187
|||||
QY 617 cccgcctgcagccttcccgagggcaccatcagcaacagagacgtatgcaagacactg 676
|||||
Db 144186 CCGCGCTGCAGGCGCTTTCCCGAGGACCATCAGCAACAGGACGATATGCAAGACACATG 144127
|||||
QY 677 tgcacgcgacagccattcagcctggccgggagcctgcctcctacagcaacgtgggtg 736
|||||
Db 144126 TGCACGCGCAGCCATTCAGCCTGGCCGGAGGCGTCCGCTCTACAGCCACGTTGGTG 144067
|||||
QY 737 ctcaaggaggtgcgtcttcttaactgcaggtgtctacccgcgtctcagcaaccccg 796
|||||
Db 144066 CTCACAGGAGGTGGCTTCTTCAACTGCAAGTGTCTTACCCGCTGCTCAAGCCCGCG 144007
|||||
QY 797 ctcaacctgcgcatcgtgcacctggtgcgcacccgcgcgcgcgtgcgtcctcccgagg 856
|||||
Db 144006 CTCACCTGCGCATGCTGCACCTGCTGCGGACCCCGCGGCGTGTGCGCTCCCGGGAG 143947
|||||
QY 857 gcgcgggcccgatactgacgcgcgcacacagcgcacatcgtgctggcaccacacgcaagtgg 916
|||||
Db 143946 GCGGCGGGCCGATATGTCACGCGCAACCGGCATCGTGTGGGCCACCAACGCAAGTGG 143887
|||||
QY 917 gtgagcgacacctcaactgcgctgattcgcagagtgctgcgagcaacgtgcgcac 976
|||||
Db 143886 GTGAGGCGCGACCTTCACCTGGCCCTGATTCGCGAGGTGTGCGGACGCCACGTGCGCATC 143827
|||||
QY 977 gcgagcgccacactcaagcgccacacctctctcgcgcgcgcgtacccgcctcgtggcg 1036
|||||
Db 143826 GCGGAGCGCGCACATCAAGCGCGCACCTTCTTCTGGCGCGCGCTACCGCTGTGGCGG 143767
|||||
QY 1037 ttgcaggaacctggcgggagcgcgtgagagatccgcgcactctacgcctcaccggc 1096
|||||
Db 143766 TTGAGGACCTGGCGGGGAGCGCGTGGCAGAGATCCGGCCTACGCTACGCTTCACCGGC 143707
|||||
QY 1097 ctgacctcacccacagctcagcctgagctgataccacaacatccacacggctcgggac 1156
|||||
Db 143706 CTGACCTCAGCGCACAGCTCGAGGCGTGGATCCCAACATCAACCCAGGCGCGGATC 143647
|||||
QY 1157 ggcaagccaatcagagccttccatacttgcctaggaatgcgcgaacgtctcccgagcc 1216
|||||
Db 143646 GSCAAGCAATCGAGCGCTTCCATCTTCTGTAGGAATGCGCGCAACGCTCTCCAGGCC 143587
|||||
QY 1217 tggcgcaagcttgcccttaactaagatctcgcgtgcagagagtgctgcgcgcgcg 1276
|||||
Db 143586 TGGCGCACGCGTGGCTTCTACATAGATCTCGGCTGCGGAGGTGTGCGCGCGCGCG 143527
|||||
QY 1277 ctgcagctcgtggctaccgctgtgactctgcggaccagcagcgtgacctcaccttg 1336
|||||
Db 143526 CTCACACTGCTGGGCTACCGGCTGTGTACTCTGCGGACACAGCGGTGACCTCACCTG 143467
|||||
QY 1337 gatctggtgtgcacagagcccgaccacttcagctgggcatcgcctgactgagaaactc 1396
|||||
Db 143466 GATCTGCTGTCGCACAGSCCCAGACCATTCAGCTGGGCATCGCTGACTGAGAACTC 143407
|||||
QY 1397 tgggctttagagcaagcccgcaactggtgcgcagcggcccgaggaagcgtgcatggtgg 1456
|||||
Db 143406 TGGGCTTTAGAGCAGCGCCCGAACTGTGTGCGCACGCGCCAGGAGCGACTGCATGCTGG 143347
|||||
QY 1457 aagagagctggggcgtatgggaa-caggtccctactatcaaccgggagtttggggctcc 1515
|||||
Db 143346 AGAGGAGCTGGGGCGCATGGGGAACAGGTCTCTACTATCAACCGGAGTTTGGGGTCC 143287
|||||
QY 1516 tccctgaagtaagcagagactgcaggtttcttctctctctgattctcgttttcttcttg 1575
|||||

Db 143286 TCCCCGAGATAGGACGACGCTGACGTTTCCTCTGATTCCTGTTTCCTTTG 143227
|||||
QY 1576 agtcttctggagctgccttctcactcaggtgcactcttctggaagcaactcttgcccc 1635
|||||
Db 143226 AGTCTCTGGAGTGCTTCTCTCATCAGGTGCATCTTTCATGG-AAAGCAACTCTTGCCCC 143168
|||||
QY 1636 taactcttctggcgagggagtaagttaactgtaataataataatgtgtgccagcc 1695
|||||
Db 143167 TACCTCTTCTGGCGCAGGAGTAAGTACTGCTAAATTAATTAATGTTGTGCCAGGCC 143108
|||||
QY 1696 ggtgcggtggctcagcctgtaatcccgacatttgaagagctgcgagcggtggatccac 1755
|||||
Db 143107 GGTGCGGTGGCTCATGCGCTGTAATCCAGCATTTTGAGAGGCTGAGCGGGTGGATCAC 143048
|||||
QY 1756 ctgaggtcagga-ttcaaaaccagcctggccaacatagtgaaaccccccctctctactaaa 1814
|||||
Db 143047 CTGAGGTCAGGAGTTCGAAACACAGCCTGGCCAACATAGTGAAACCCCTCTCTACTAAAA 142988
|||||
QY 1815 atgcaaaattagtcggtgcgtgcacactcctgtaatcccgactacttagagctg 1874
|||||
Db 142987 ATCAAAATTAGTCCGCGGTGTGGCACACTCTCTGTAATCCAGCTACTTAGAGGCTG 142928
|||||
QY 1875 agtgggaaatacaacttgactccaaaggtggaggttgcagtaagctgaaatcatgcac 1934
|||||
Db 142927 AGTGGGAGAAATCACCTGGACTCCAGAGTGGAGTTGCAAGTGAAGCTGACATGCGCAC 142868
|||||
QY 1935 tgacacctagctgggtggcaagcaaacactctatacaaaaaataataataattgtt 1994
|||||
Db 142867 TGACCTTAGCTGGGTGGCAGAGCAAGACTATCAAAAAATAAGTAATAAATTTGTT 142808
|||||
QY 1995 caaaagtctctgc 2006
|||||
Db 142807 CAAAGTCTGTC 142796
|||||
RESULT 4
AC025287 194832 bp DNA linear PRI 26-JAN-2002
LOCUS Homo sapiens chromosome 16 clone RP11-490B18, complete sequence.
DEFINITION AC025287
ACCESSION AC025287.8 GI:18376863
VERSION HTG.
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE 4 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On Jan 26, 2002 this sequence version replaced gi:17976465.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.4% of Sequence;
Estimated Total Number of Errors is 0.6.

[illegible]


```
QY 1097 ctgacctcaccgcaagctcgagcctggatccacaacatcacccacgggtcggggac 1156
Db 62279 CTGACCTCTACGCCACAGCTCGAGGCTGGATCCACAACATCACCCACGGTTCGGGGATC 62338
QY 1157 ggcagccaatcgagcccttcatacttcgtctaggaatgcgcaacgctctccaggcc 1216
Db 62339 GGCAGGCCAAATCGAGGCTCTCATCTTCGCTAGGAATGCGCAACGCTCTCCAGGCC 62398
QY 1217 tggcgcaagcgttgccctcactaaagatcctgcgctgagggagtgctgcggcgccg 1276
Db 62399 TGGCGCCACGGTTCGCCCTTCACTAAGATCTGCGCGTCGAGAGGTGCGCCGGCCGG 62458
QY 1277 ctgcagctctgggtacccgctctgtgtactctgagaccagcagcgtgaacctaccctg 1336
Db 62459 CTGCAGCTCTGCGGTACCGGCTGTGTACTCTGGGACCAAGCAGCGTGAACCTCACCCCTG 62518
QY 1337 gatctgtgtgccacagggcccgaccacttaagcttggtgcatcgctgactgaaactc 1396
Db 62519 GATCTGTCGTGCCACGAGGCCAGACCACTTCAGCTGGGCACTGCGCTGACTGAGAACTC 62578
QY 1397 tgggccttagagcaagcccgaaactgtgtgcccaggcccgaggaagcgaactggtgg 1456
Db 62579 TGGGCTTAGAGCAGGCCCGGAACCTGTGTGCCAGGCCCGAGGCGACTGCATGGTGG 62638
QY 1457 aaaaggagctggggcgcatgggaa-caggtccctactatcaaacgggagtttgggtcc 1515
Db 62639 AGAGGAGCTGGGGCGCATGGGGAAGCAGGTCCTACTATCAACCGGAGTTGGGGTCC 62698
QY 1516 t-ccccgaagaaagcaagactcacgcttcttctctctgattctcgttttctctt 1574
Db 62699 TCCCCCTGAAGTAGCAAGGACTGCAGTTCCTTCTCTCTGATCTCGGTTTCTCCCTT 62758
QY 1575 gagtctctggagctgcctctctatcatcaggtgcactcttctggaagcaactcttggcc 1634
Db 62759 GAGTCTTCTGGAGTGCTCTCTCATCAGGTGCACCTTCATGG-AAAGCAACTCTTGCCC 62817
QY 1635 ctacctcttgccgagcaggagtaagtactgctaaataataatgctgctccaggc 1694
Db 62818 CTACCTCTTCTGGCGCAGGAGTAAGTTACTGTAAATTAATTAATGTGTGCCAGGC 62877
QY 1695 cgggtcggtggctcatgctgtaatcccagcatttgagagcgtgagggcggtggatca 1754
Db 62878 CGGTCGCGTGGCTCATGCTGTAATCCAGCACTTTCAGAGGCTGAGCGGGTGGATCA 62937
QY 1755 cctgaggtcagga-ttcaaacacagcctggccaaacatagtgaaacccctctctactaaa 1813
Db 62938 CCTGAGTCAAGGAGTTGCAAAACAGCCTGGCCAAACATAGTGAACCCCTCTCTACTAAA 62997
QY 1814 aatgcaaaaattagtcggcgctggtgacacactctgtaatcccagctacttagagcgt 1873
Db 62998 AATGCAAAAATTAGTCGGCGGTGGTGGCACACTCTCTGTATTCACAGCTACTAGGAGCT 63057
QY 1874 gagggtggaaaatacacttgactcccaaggtggaggtgagtaagctgaaatcatgccca 1933
Db 63058 GAGGTGGGAGAAATCACTTGGACTCCAGAGTGGAGGTTGCAGTAAGCTGAGATCATGCCA 63117
QY 1934 ctgcacctagcttggtggaagcaaaactctatcaaaaaataataataatttgt 1993
Db 63118 CTGCACCTAGCTTGGTGGTGGCAGCAGAGACTCTATCAAAAAATAAGTAATAAATTGT 63177
QY 1994 tcaaaagtcctgc 2006
Db 63178 TCAAAAGTCTGC 63190

RESULT 6
AF219991 71503 bp DNA linear PRI 26-OCT-2000
LOCUS AF219991
DEFINITION Homo sapiens intestinal N-acetylglucosamine-6-O-sulfotransferase
(CHST5) and corneal N-acetylglucosamine-6-O-sulfotransferase
(CHST6) genes, complete cds.
ACCESSION AF219991
VERSION AF219991.1 GI:11023147
```

```
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 71503)
AUTHORS Nakamura,T., Dotsu,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S.,
Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.
Maternal corneal dystrophy type I and type II are caused by distinct
mutations in a new sulphotransferase gene
Nat. Genet. 26 (2), 237-241 (2000)
TITLE
JOURNAL
MEDLINE 20472330
PUBMED 11017086
REFERENCE 2 (bases 1 to 71503)
AUTHORS Akama,T.O. and Fukuda,M.N.
TITLE Direct Submission
JOURNAL Submitted (29-DEC-1999) Glycobiology Program, The Burnham
Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
source
1. 71503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22"
/clone="CITB-483K2"
358..461
/rpt_family="L2"
/rpt_type-dispersed
repeat_region
complement(462..771)
/rpt_family="Alu"
/rpt_type-dispersed
repeat_region
772..945
/rpt_family="L2"
/rpt_type-dispersed
repeat_region
1169..1338
/rpt_family="MIR"
/rpt_type-dispersed
repeat_region
complement(1533..1804)
/rpt_family="Alu"
/rpt_type-dispersed
repeat_region
2616..2916
/rpt_family="Alu"
/rpt_type-dispersed
misc_feature
3309..3532
/note="similar to EST H12297 (EST cluster Hs.31147)"
misc_feature
3865..4028
/note="similar to EST H12297 (EST cluster Hs.31147)"
repeat_region
4099..4406
/rpt_family="Alu"
/rpt_type-dispersed
repeat_region
4793..5250
/rpt_family="L2"
/rpt_type-dispersed
repeat_region
complement(5251..5329)
/rpt_family="MIR"
/rpt_type-dispersed
repeat_region
complement(5768..6080)
/rpt_family="Alu"
/rpt_type-dispersed
repeat_region
complement(6081..6265)
/rpt_family="Alu"
/rpt_type-dispersed
misc_feature
6111..6567
/note="similar to EST AA001593 (EST cluster Hs.6853)"
repeat_region
8035..8350
/rpt_family="Alu"
/rpt_type-dispersed
misc_feature
10545..11049
/note="similar to EST AI950923 (EST cluster Hs.6853)"
repeat_region
11043..11195
/rpt_family="Alu"
/rpt_type-dispersed
```



```

/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="BAD56001.1"
/db_xref="GI:5917708"
/translation="MWLPFRSSKTVVLLIAQTCLLLFIILSRPSPSPAGGEDRVHV
LVLSMRSGFLSOLFQSDHFDYLMPEAWHVTLSQGSAAATHMAVRDLMSIFL
CDMDVDFAYMPSQSLNLSAFENWATSRALCSPPAGSAFPRGTISKQDVCKTLCTQPFS
LAREACSYSHVILKEVREFNLQVPLLSPALNRIVLHVRDPRPRAVLKREAAAGPI
LARDNGVLVTGNGKWRVDPHRLILREVCSHRVLAEAATLKPPPLRGVRLVRFED
LAREPLAERIALYAFITGLTLPQLEAWHTNTHSGSGIKPKPIEAHTSSRNARNYSQAW
KHALFFPKILRVQVRCAGAILQLLGYRYSADQDQDLTLDLVLPRGPHFWSWSPD"
BASE COUNT      239 a      510 c      446 g      267 t
ORIGIN
Query Match      61.1%; Score 1248.6; DB 9; Length 1462;
Best Local Similarity 99.7%; Pred. No. 2.3e-188;
Matches 1251; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 197 cgtgtgccccagccgcgcagctgtggctggccacaggttctccagcaagacagtcagtg 256
Db 208 CGAAGGCCCCCGAGCGCCCGCAIGTGGCTGCCACAGGTTCTCCAGCAAGACAGTGACAGTG 267
QY 257 ctctctggcagacagaccctgcctctgtctctcatctcccgccagggccctca 316
Db 268 CTCTCTCTGGCACAGACACCTGCTCTGCTCTTCATCATCTCCGGCCAGGGCCCTCA 327
QY 317 tcccagccggcgagagatcgtgtgacagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 376
Db 328 TCCCCAGCCGGCGGCGAGGATCGTGTGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 387
QY 377 tcatctcttggccagctcttccagcagccccagcagctcttctacctgatggagccc 436
Db 388 TCATCTCTCTGGCCAGCTCTTCAGCAGACCCCGAGCTCTTCACCTGATGGAGCCC 447
QY 437 ggtgtgcatgtgtggaccacctgttcgaggcgagcgcgagcgagcgagcgagcgagcgag 496
Db 448 GCGTGCATGTGTGGACCACTGTTCGACGGCGAGCGCGAGCGCGACGCTGCGCTGTGCGCTG 507
QY 497 cgcgacctatgcctctatcttttgcgacatgagcagctgtttgatccatcacatgcca 556
Db 508 CGGACCTGATGGCTGCTATCTTTTGTGCGATGAGCTGTTGTGCTATGCTATGCTATGCTCA 567
QY 557 cagagccgaacctgtccgcctttttcaactggcagcagcgagcgcgctgtgtctgcgcg 616
Db 568 CAGAGCCGAAACCTGTCCGCTTTTCACTGGCAAGACGCGCGCGCTGTCTCTCGCG 627
QY 617 ccgcctgcagcgcttcccgaggacacatcagcaagcagcagcagcagcagcagcagcagc 676
Db 628 CCGCCTGCAGCGCTTCCCGAGGACACATCAGCAAGCAGGACGATGCAAGACACTG 687
QY 677 tgacgcgagcattcagctgtcccgaggcgctgcctcctacagccagcagcagcagcagc 736
Db 688 TGACCGCGGAGCATTCAGCTGTGCGCGGAGGCGCTGCGCTCTTACAGCCAGTGGTG 747
QY 737 ctaagagtgcgcttcttcaacctgcaggtgtctctacccgtgtctcagcgaccgccg 796
Db 748 CTCAGAGGTGGCTTCTCAACCTGCAGGTGCTCTACCGCTGTCTACGCGACCCCGCG 807
QY 797 ctaacctgcgcatcgtgacctgtgtgcgagaccgcggcgcgctgtgtgtgtgtgtgtgtgt 856
Db 808 CTCAACTGCGCATCTGCACTGTGTGCGGAGCCGCGGCGCGTGTGTGTGTGTGTGTGTGTGT 867
QY 857 gcgcgggcccgatctggcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 916
Db 868 GCGGCGGCGCGGATGCGGACCGGACCAACCGGATCTGTGTGTGCGACCAACGCAAGTGG 927
QY 917 gtgagggcgacacctcactgcctgatttcgaggtgtgtgcgagccagcagcagcagcagc 976
Db 928 GTGAGGCGCGACCTTCACTTGGCGCTGATTTCGCGAGTGTGTGCGGACGCGCGCGCATC 987
QY 977 gccgagggcgcaactcaagcgccacaccttctgtcgcgcgcgctaccgctgtgtgtgtgtgt 1036
Db 988 GCGGAGCGCGCACACTCAAGCGCGCACCTTCTGTGCGCGCGCGCTACCGCTGTGTGTGTGT 1047

```

```

QY 1037 ttccaggacctgcgcgggagccgctggcagagatccgcgcactctaccccttccacggc 1096
Db 1048 TTCGAGGACCTGGCGGGAGCGCGCTGGGAGAGATCCGGCGACICTACCGCTTACCGCGC 1107
QY 1097 ctgacctcacgccacagctcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtcagcgtc 1156
Db 1108 CTGACCTTACGCCACAGCTCGAGGCTGGATGCCAACATCACCCACGGTGGGGATC 1167
QY 1157 gcaagccaatcagcgtccttcatactctgtctagaaagcgcgcaagcttctcccaaggcc 1216
Db 1168 GGCAGGCCAATGAGGCTTCCATCTTCTGTAGAAATGCGCGCAACGCTCTCCCGAGGCC 1227
QY 1217 ttgcyccacgcttgcccttccactaagatccctgcgctgcagagagtgctgcgcggcgcg 1276
Db 1228 TGGCGCCACGCGTGGCTTCTACTAAGATCCTGCGGTGCGAGAGGTGTGCGCGCGCGCG 1287
QY 1277 ctgcagctcgtggctaccggcctgtgtactctgcggaccagcagcagcagcagcagcagc 1336
Db 1288 CTGCAGCTGCTGGCTTACCGGCTGTGTACTCTGCGGACCAGCAGCGTGACCTCACCTG 1347
QY 1337 gatctggtgctgcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1396
Db 1348 GATCTGCTGTGCTGCGAGGCGCGACACCTTCAGCTGGGCAATCGCTGACTGACTGAGAATC 1407
QY 1397 tggccttagagcaagcccgcaactgtgtgcgagggcccgaggaagcagcagcagcagc 1451
Db 1408 TGGGCTTAGAGCAGGCGCGCGCACTGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 1462
RESULT      8
AF280086    3786 bp    mRNA    linear    PRI 20-FEB-2001
LOCUS      Homo sapiens N-acetylglucosamine 6-O-sulfotransferase GST-4beta
DEFINITION mRNA, complete cds.
ACCESSION AF280086
VERSION    AF280086.1 GI:12060803
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 3786)
            Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
            Rosen,S.D.
            Chromosomal localization and genomic organization for the
            galactose/ N-acetylglactosamine/N-acetylglucosamine
            6-O-sulfotransferase gene family
            Glycobiology 11 (1), 75-87 (2001)
            21095027
REFERENCE  2 (bases 1 to 3786)
            Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and
            Rosen,S.D.
            Direct Submission
            Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
            3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES   Location/Qualifiers
            source          1..3786
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /map="16q23.1"
            gene            1..3786
                        /gene="GST4beta"
                        /note="clusters with the intestinal N-acetylglucosamine
                        6-O-sulfotransferase (GST4alpha) gene in GenBank Accession
                        Numbers AF176838 and AF176839"
            CDS              189..1376
                        /gene="GST4beta"
                        /note="similar to N-acetylglucosamine 6-O-sulfotransferase
                        GST-4alpha"
                        /codon_start=1
                        /product="N-acetylglucosamine 6-O-sulfotransferase"

```

[illegible]

mutations in a new sulphotransferase gene
 Nat. genet. 26 (2), 237-241 (2000)
 11017086
 2 (bases 1 to 2544)
 Akama,T.O. and Fukuda,M.N.
 Direct Submission
 Submitted (23-DEC-1999) Glycobiology Program, The Burnham
 Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA
 Location/Qualifiers
 1. .2544
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="16"
 /map="16q22"
 1. .2544
 /gene="CHST6"
 /note="responsible for macular corneal dystrophy"
 693..1880
 /gene="CHST6"
 /function="transfers sulfate to 6-OH of GlcNAc in keratan
 sulfate"
 /note="C-GlcNAc6St"
 /codon_start=1
 /product="corneal
 N-acetylglucosamine-6-O-sulfotransferase"
 /protein_id="AAG26325.1"
 /db_xref="GI:11023146"
 /translation="MMLPRVSSVAVTALLIAQTFLILFLVSRPGPSSPAGEARVHVL
 VLMSWRSGSFQVLEQNHQDFVLMPEPMEVHTLTSQSSAAHLHVAVRDLVRSVFLL
 DMVEFDAYLPWRNLSLDLQWVYSRALCSPACSAUPRGAISSEAVCKPCAROSFTPL
 ARACRSYSHVILKEVRFNQLVDPLLSDFALNRIVHLVRDPRAVLRREGTAKAL
 ARONGIVLGTNWADPGLRVREVRVREKSHVRIAEATLKPFPFLRGVRLVRFDEL
 ARPELAIRALYFTGLSLTPQLEAWIHTHSGSPCARREAEKTSRNALNYSQAWR
 HALFPKIRRVQELCAGALLQLGYPVYSEDEQRNALDLVLRPLNGFTWASSIASH
 PRN"
 BASE COUNT 460 a 799 c 733 g 552 t
 ORIGIN
 Query Match 51.5%; Score 1052; DB 9; Length 2544;
 Best Local Similarity 85.0%; Pred. No. 2,8e-157;
 Matches 1240; Conservative 0; Mismatches 200; Indels 18; Gaps 5;
 QY 140 tcgagacatctccatggattgtgtatcagcgttgtagctctccacagccctggcgg 199
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 615 TCTAATCATCTCCATGGATTGTGATCAGCGTTGCAGCTCCACGAGCCCTGGACAG 674
 QY 200 tggcccccagcggcccgatgtgctccacggttctccagcaagacagtgacagtctc 259
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 675 TGGCCCCCAGCAGTCACATGTGGCTCCGCGCGCTCTCCACACACAGCAGCAGCGCTC 734
 QY 260 ctctggcacagacacctgcctctgtctctcatcatctcccgccagggccctcatcc 319
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 735 CTCCTGGCGCAG---ACCTTCCCTCTCTCTTCTGTGTTTCCCGCCAGGCCCTCTGTC 791
 QY 320 ccagccggggcgaggatcgtgcaagctggtgtgctgctctgtgctgctggcgctca 379
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 792 CCAGCAGCGCGCAGGCGCGCTGCTATGTGTGTGTGCTCTCTGTCCTCGTGGCGCTCGGCGTCG 851
 QY 380 tctctctggccagctctcttcacccagcaaccgcagcgtctcttacctgattggagccgcg 439
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 852 TCCCTGTGGGCGCAACTCTTCAACACGACACCCGACGCTCTTACCTAATGAGACCGCG 911
 QY 440 tggcatgtggacacacccctgtcgagggcagcgcgcaacgctgcaatggcgtgcgc 499
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 912 TGGACAGTGTGGACACCCCTGTGCGAGGCGAGCGCCGCAACGCTGCACATGGGTGTGCGC 971
 QY 500 gaactgatgcgtctatctcttttggcagatggacgtgtttgatgctactatgccacag 559
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 972 GACCTGTGGCGCTCGGCTCTCCCTGTGCGACATGGACGTGTTGTAGTCCCTATCTGCCCTGG 1031
 QY 560 agccgaaactgtccgcctttttcaactggggcaacgagcgcgctgtgctgcgcgcgc 619

[illegible]

Query	Subject	Score	Length	Identical	Similarity	Mismatches	Indels	Gaps
Db 108615	TGTTGCTGCCACGAGGCTGAACGGCTTCACWTGGGCAATATCCACCGCTCGCACCC	46.3%	947	DB 6	Length 1647			
Qy 1399	ggccttagagcaagcccgaaactgtg--gtcgccaggcccggaagcgaactgcattgtgg	85.1%	Pred. No. 1.1e-140					
Db 108675	GAAATAGTGAGGCCACAGTGTAGCAGCGCTAGGCCCGGAGGAGTGTGCTGATC	0	Mismatches 185					
Qy 1457	aaaagagctggggcgatgggaa--caggtccctactatcatcaacgggagtttgggt--							
Db 108735	AGAGGGGCTGGGGCGCACCGGAGAACAGGTCCCTATATGACCAAGAGTTCGTGTAC							
Qy 1514	-----cctccctgaagtaagaagaaactgcagcttttctctctgattctcgg							
Db 108795	GACCCCTCCCCCTCCCAAGTAGCAAGGATGCACGTTCTTCTCTCTTGTGATCTTGG							
Qy 1566	ttttctttgactctttggagctgcctcttcatacagtgacactcttctcatgaaaagcaa							
Db 108855	TTTTCTTTGAGTCTCTGGAGCTGCCCTCTCATCAGTGCACCTTCATCG--AAAGCAA							
Qy 1626	ctctggccctacactcttctgggagcagggagtaagt 1663							
Db 108914	CTCTTGGCCCT--CTTCCCTTGGGCACAGGGTGTGGCT 108950							
RESULT 12								
LOCUS	AX327330	1647 bp	DNA	linear	PAT 07-JAN-2002			
DEFINITION	Sequence 15 from Patent WO0179468.							
ACCESSION	AX327330							
VERSION	AX327330.1	GI:18097876						
KEYWORDS	human.							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
	1 (sites)							
	Policky,J.L., Hafalia,A., Burford,N., Ring,H.Z., Lal,P.,							
	Triboulet,C.M., Yao,M.G., Yue,H., Tang,Y.T., Patterson,C., Das,D.,							
	Sanjanwal,S.M., Gandhi,A.R., Reddy,R., Khan,F.A., Baughn,M.R.,							
	Ramkumar,J., Griffin,J.A. and Au-Young,J.							
TITLE	Drug metabolizing enzymes							
JOURNAL	Patent: WO 0179468-A 15 25-OCT-2001;							
FEATURES	Incyte Genomics, Inc. (US)							
source	1..1647							
	/organism="Homo sapiens"							
	/db_xref="taxon:9606"							
BASE COUNT	287 a 560 c 501 g 299 t							
ORIGIN	/note="Incyte ID No: 7472777CBI"							
Query Match	46.3%; Score 947; DB 6; Length 1647;							
Best Local Similarity	85.1%; Pred. No. 1.1e-140;							
Matches 1106;	Conservative 0; Mismatches 185; Indels 8; Gaps 4;							
Qy 218	atgtggtgccaggtttctccagcaagacagtgacagtgctctctctggcacagaccacc	277						
Db 1	ANGTGGCTCGCGGGCTCTCCAGCACAGCAGTACCGGCTCTCTCTGGCGCAG--ACC	57						
Qy 278	tgctctctcttctatcatctccggccagggccctatcccccagcggcgagat	337						
Db 58	TTCTCTCTCTCTTTCTGTTTCCCGGCCAGGGCCCTGTCCTCCACAGCGCGGAGCG	117						
Qy 338	catgtacagctagtgctctctcgtagcgtcggtctcatctcttggccagctc	397						
Db 118	CGCGTCATGTGTGGTGTCTCTCTGTGGCGCTCTGGGCTCGTCTCTGTGGCCAACTC	177						
Qy 398	ttcagccagcaccgcagctcttctacotgatggagcccgctggcatgtgtggaccacc	457						
Db 178	TTCAACACGACCCGACGCTCTTACCTAATGGAGCCCGCTGGGACCGCTGTGGACACC	237						
Qy 458	ctgtgcagggcagcggcgaagctgcacatggccgtgcgcgaactgatgcgtctatc	517						

[illegible]

RESULT 13

```

AF176840
LOCUS       AF176840               1989 bp    mRNA    linear    ROD 22-SEP-1999
DEFINITION Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase
            (I-GlcNAC-6-ST) mRNA, complete cds.
ACCESSION   AF176840
VERSION     AF176840.1  GI:5917709
KEYWORDS    house mouse.
SOURCE      Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 1989)
AUTHORS    Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
TITLE      Cloning and characterization of a mammalian
            N-acetylglucosamine-6-sulfotransferase that is highly restricted to
            intestinal tissue
JOURNAL    Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE    99423499
AUTHORS    Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
TITLE      Direct Submission
JOURNAL    Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience,
            3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES   Location/Qualifiers
            1..1989
             /organism="Mus musculus"
             /strain="C57Bl/6"
             /db_xref="taxon:10090"
             /chromosome="8"
             /map="8E1"
             /tissue_type="intestine"
            1..1989
             /gene="I-GlcNAC-6-ST"
             107..1294
             /gene="I-GlcNAC-6-ST"
             /note="possibly involved in biosynthesis of mucosal
             ligands for I-selectin"
             /codon_start=1
             /product="N-acetylglucosamine 6-O-sulfotransferase"
             /protein_id="AA056002.1"
             /db_xref="GI:5917710"
             /translation="MLRPFSSVTMLSLLMVQTHILFLVSROVPSSPAGLGRVHVL
             VLSWSGGSFVQLSQHPDLYMLPEAHVWDTLSQGSAPALHNAVRLDLSVLC
             DMVFDAYLPWRNISDLFWAYSRALCSPPVCEAFARGNISSEVCKPLCAIRPGL
             AQACSSIVSHVZKVEFFNLQYELISDPALNRIVHLVDRPRAVLRSREGTAKAL
             ARDLPIVIRELYATFGLTGLTPQLTQIHNTTSGSGPARAEAFKTTSDALSVSQWAR
             HLTPLFAKIRRVQELCGALQLLYGRSVHSELEQRDLSLDLLPRGMDSEKASSTEQ
             PES"
BASE COUNT 415 a 586 g 550 g 438 t
ORIGIN
Query Match 36.8%; Score 751.2; DB 10; Length 1989;
Best Local Similarity 70.9%; Pred. No. 1e-109;
Matches 1046; Conservative 0; Mismatches 413; Indels 17; Gaps 3;
QY 202 gccccacgcccgcgagtgaggctgcacggtctccgcagcaacagtgacagtgtctcct 261
    ||| |||||
Db 91 GGCCCTATCCAGAGAGGTATGCGGTATCCCGCTTTCTCCGACACTGTCTATGCTTTCGCTCT 150
QY 262 cctggcacagacacctgcctcctgctcttcacatcctccggccagggccctcacc 321
    ||| |||||
Db 151 GATGTACAG--ACYGGCATCTGGTCTTCCTGTGCTCCCGGCAAGTGCATCGTCC 207
QY 322 agcggcgggcgagdgatggtgcacgtggtggtgctcctcgtggcgctcggggtcacc 381
    ||| |||||
Db 208 AGCAGGCGTCTGGGAGCGGTGTGCAGGTGCGGTACTGTCTCTGTCGGCTCGGGCTCG 267
QY 382 ctctctggccagctcttcacagcagaccccgacgtctcttacctgatggagcccgctg 441
    ||| |||||
Db 268 CTTCTGTGGGCAGCTCTACGCAACACCCGATGCTTCTACCTCATGAGCGCGCTG 327

```

QY	442	gcattgttgaccacccctgtctcgacgggcagcgcggaacacgctgacatggcctgctgcgca	501
Db	328	GCAGCTCTGGGATATCATGTTGTCCGACGGGAGTGCOCOCACATCCACATGGCCGTGGTGA	387
QY	502	cctgatcgctctatcttttctgcacatggacgtgttctgatgctacatgcccacagag	561
Db	388	CCTGATCCGTCAGTGTTCCTATGGACATGGACGTATTGTATGGCTACCTGCCCTGGCG	447
QY	562	ccgaacctgtccgccttttctaacctgggcaacgagcgcgcgtgtgtctgcgcgcgcgc	621
Db	448	CCGCAACATCTCGGATCTCTTCCAGTGGCGGTGAGCGCGCATTTGTCTCACTCCCGGT	507
QY	622	ctgcagcgcttttcccgcagcaccatcagaacagcagacgtatccaagacactgtgac	681
Db	508	CTGGAAACCTTCGTCGTGTGCAACATCAGACAGCAGGAGGTGTGTAACTCTGTGCGC	567
QY	682	gcggcagcattcagcctggccggagagcctgcgcgtctctacagccacgtggtctca	741
Db	568	AACGCGGCCCTCGGCCCTGAGCAAGCCCTGACAGAGCCTCTATATGTCAGTGGTCTCAA	627
QY	742	ggaggtgcgctcttcaacctgcaggttactctaccgcgtctcagcaccgcgcgctca	801
Db	628	GGAGTGGCGCTCTTTTAACTACAGGTGCTCTACCGCTGCTCAGCGACCTTCGGCTCAA	687
QY	802	cctgcgcatcgtgcacctggtgcgcacgcgcgcgcgtgctgcgtcccccggagcgcgc	861
Db	688	CCTGGCATGTGTGCACCTAGTSCGCGACCCGCGGCCGTGCTGCGCTCCCGAGAGCAGAC	747
QY	862	gggcgcgatactgcagcgcgaacacggcatcgtgtgtggcacaacacggcaagtgggtga	921
Db	748	AGCCAAAGCGCTTGGCAGCGGACATTTGGATCTCTCTGGGTACCAAGCGCACCTGGTGG	807
QY	922	ggcgcacctcacctgcgcctgtatcgcgaggtgtgcgcgacgcacgtgcgcatcgcca	981
Db	808	GSCGACCCCGGCTGCGGTGTCAACGAGGTATGCCGACCCATGTGCGCATCTCAGAC	867
QY	982	ggcgcgcacactcaagcgcgcacctctctgcgcgcgcgtaccgcctggtgcgcttga	1041
Db	868	GGCAGCCTTGCACAGCGCGCCCTCTTGAAGATCGCTACCGCTGGTGGCTGACGAC	937
QY	1042	ggacctgcgcgggagcgcgtgcgagatccgcgcactctacgctctcacgcgcctgac	1101
Db	928	GGATCTGGCCGGGACCCACTCACCGTAACTCGTGACTCTATGCTTACCGGCCCTGGG	987
QY	1102	ctctacgcacagctctgagcctggatccacaacatcccccgggtcggggatcgga	1161
Db	988	TCCTACGCGGAGCTCCAGACTTGGATCCACAATATCAGCATGTGTCAAGGCGACGCGC	1047
QY	1162	gccaatcgaagccttcataacttctctaggaatcgcgcaacgctccaggtcctggcg	1221
Db	1048	GGCCCGTGAAGCCTTCAAGCACATCCAGGATGGCTCAGTGTATCCAGGCCCTGGCG	1107
QY	1222	ccacgcttgcccttcaataagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgtgca	1281
Db	1108	CCACACGCTGCCCTTTGCCAAGATCCCGGCTCCAGAACTGTGCGGGGGTGCACCTGCA	1167
QY	1282	gctgtgggtcaccgctctgtactctctgcgacacagcgcgtgacctcacccctgatct	1341
Db	1168	GCTGCTGGGTWACCGGCTCTGCATTCGGAGCTTGAGCAAAAGGACCTCTCTTGGACCT	1227
QY	1342	ggtgctgcacagggcccagacacttcagctgggcgcgcgtgactgagaaactctgggc	1401
Db	1228	CCTGCTGCCAAGAGCATGGACAGTTTCAAGTGGGATCTGTCACCGGAGNAGCAACCGGA	1287
QY	1402	cttagagcaagcccgaactgtggtgcgcagggcccaggaagcgcactgactggttgaaaag	1461
Db	1288	ATCTTAGAATTTT---AGTGAGAGACCCACGCTATAAATATTAGGGTCTATTGGAGTATGA	1344
QY	1462	gagctggggcgcatgggaacaggtccctactatcaacgcgggaatttggggtcctccct	1521
Db	1345	TAAAGAGGGGCTTGGAGAAACCCAAAGCAAGTAGCTGGGAGTGTGAGTGATCTTCTCCT	1404
QY	1522	gaagtaagcaagactg-----cacgtttctctctctgactccaggttttc	1570

Db	1405	GTACTAGAAAGAGTGGAGTCAAAATCCCATCTCTTCTGTCAGATTTGAGTTTCG	1464
Qy	1571	ctttgagtccttcgagtcgcttcctcctcatcaggtcgacotcttcattcgaaagcaactott	1630
Db	1465	GTGGTGCTTTTAGGGTTGGATTCCTCCACCAAGTACTATCGAATGGAAGCAAAAGCTGT	1524
Qy	1631	gccctacactcttcggcgaggaggaagttaact	1666
Db	1523	GCCCACTTCTCCAGAGAGGAGCCAGCCCTCTACT	1560
RESULT 14			
AF176841			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
mrna			
gene			
cds			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches 920; Conservative			
36.7%; Score 750.4; DB 10; Length 1740;			
Pred. No. 1.4e-109;			
Indels 3; Gaps 17;			

Query Match 36.7%; Score 750.4; DB 10; Length 1740;
Best Local Similarity 77.7%; Pred. No. 1.4e-109;
Matches 920; Conservative 0; Mismatches 261; Indels 3;

Search completed: September 24, 2002, 14:16:59
Job time: 29258 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 09:23:44 ; Search time 10470.7 Seconds
(without alignments)
2634.765 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggcctcaggtccactgtct.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*
2: em_esthum.*
3: em_estm.*
4: em_estm.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_esti.*
10: gb_est2.*
11: gb_hic.*
12: gb_gss.*
13: em_gss.hum.*
14: em_gss_inv.*
15: em_gss_pin.*
16: em_gss_vrt.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	657.4	32.2	695	10	BE857538
C 2	631	30.9	735	10	BE858652
C 3	525	25.7	525	10	BF197521
C 4	487.6	23.9	620	9	AI824100
C 5	420.8	20.6	436	9	AW081348
C 6	352.4	17.2	1923	11	AK009113
C 7	346.4	16.9	1067	12	CNS03KN7
C 8	322.8	15.8	657	12	AG035205
C 9	308.4	15.1	517	10	BF042384
C 10	285.4	14.5	568	9	AI115260
C 11	284.2	13.9	954	10	BI823850
C 12	279.4	13.7	849	12	CNS04QFN
C 13	260	12.7	260	10	BF824736
C 14	259.4	12.7	955	10	BI107354
C 15	245.4	12.0	433	12	A2405100
C 16	244.6	12.0	429	9	BB49113
C 17	231.4	11.3	337	10	BE960153

18	224.4	11.0	595	10	BG964671
19	224.4	11.0	783	10	BG963298
20	218.8	10.7	852	10	BG966340
21	217.4	10.6	624	12	AQ373399
22	214.2	10.5	1655	11	BG022315
C 23	213.8	10.5	560	9	AL589348
24	211.4	10.3	452	10	BF225761
25	211	10.3	352	9	AW089016
26	210.4	10.3	515	10	BM467754
27	210.2	10.3	400	9	AV760391
28	210	10.3	388	12	AQ557029
C 29	209.4	10.2	693	12	AG118065
C 30	209.2	10.2	412	9	BG062478
31	209	10.2	539	12	AQ379787
32	208.4	10.2	1172	11	AF305824
C 33	208.2	10.2	416	9	AW341978
34	208	10.2	2048	11	BC017051
35	207.8	10.2	616	12	AQ554309
C 36	207.2	10.1	451	9	AV695478
C 37	207.2	10.1	650	12	AQ530872
C 38	207	10.1	384	9	AL040054
39	207	10.1	946	10	BG335756
C 40	206.6	10.1	602	12	AQ540860
C 41	206.4	10.1	345	10	BG222875
42	206	10.1	544	10	BG739841
43	206	10.1	544	12	AQ394650
C 44	205.8	10.1	688	12	AG118999
45	205.6	10.1	440	10	BG189911

ALIGNMENTS

RESULT 1
BE857538/c

LOCUS

DEFINITION

ACCSSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

BE857538 695 bp mRNA linear EST 29-SEP-2000
7g01a08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3'
similar to TR:Q9UBX3 Q9UBX3 N-ACETYLGLOUCOSAMINE
6-O-SULFOTRANSFERASE.; mRNA sequence.

BE857538

BE857538.1 GI:10371664

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1. (bases 1 to 695)

NCI/MINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTCAP), Tumor Gene Index

Unpublished (1998)

Contact: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -400P from Gibco

High quality sequence stop: 431.

Location/Qualifiers

1..695

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3305174"

/clone_lib="NCI_CGAP_Brn23"

/tissue_type="glioblastoma (pooled)"

/lab_host="DH10B"

/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 127 a 210 c 247 g 111 t
ORIGIN

Query Match 32.2%; Score 657.4; DB 10; Length 695;
Best Local Similarity 97.6%; Pred. No. 1e-67; 16; Indels 1; Gaps 1;
Matches 678; Conservative 0; Mismatches 16;

QY 284 ctgctcttcattatctccggcagggccctcatcccccggcgccgagagatcggtg 343
|||||
Db 695 CTGCTCTTCATCTGCGCGCAGGGCCATCATCACCGCCAGCGCATCGTGTG 636
|||||
QY 344 caagtctgctgctctctggtgctcggtctcctctctctctctctctctcagc 403
|||||
Db 635 CAGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 577
|||||
QY 404 cagcaccgagctctctacatgatggagccgctggtgcatgtgtggaccacccctgctg 463
|||||
Db 576 CAGCACACGACGCTCTTCTACTTGTATGAGGAGCGCGGTGATGTGGACCATCTGTCG 517
|||||
QY 464 caggcgagcgcggaacgtgacatggcgtgctgctgctgctgctgctgcttctt 523
|||||
Db 516 CAGGCGAGCGCGCAACGCTGCACATGCGGTGCGGACGTGATGCTATCTTTTG 457
|||||
QY 524 tggacatgagctgtttatgctctacatccacagccgagcgaacctgtccgcttttc 583
|||||
Db 456 TGGAGATGACGCTGTGTAIGCTTACATGCCACAGAGCCGAACCTGTCCGCTATTTC 397
|||||
QY 584 aactgggcaagcagcgcgctgtgctgctgctgctgctgctgctgcttccccgagc 643
|||||
Db 396 AACTGGGCAACGAGCGCGCGCTGTGCTGCGCGCGCGCTGCGAGCGCTTTCGCCGAGGC 337
|||||
QY 644 accatcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 703
|||||
Db 336 ACCATCAGCAAGCAGGACGCTATCAAGACACTGTGACGCGGCGGACCATTCAGCGCTGCC 277
|||||
QY 704 cggagcctgctgctctcctacagccagctgtgtctcagagagtgctctctcctc 763
|||||
Db 276 CGGAGGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 217
|||||
QY 764 caggtgctctaccgctgctcagaccccgctcagaccccgctcagaccccgctcagaccccgctg 823
|||||
Db 216 CAGGTGCTCTACCGGCTGCTCAGCGACCGCGCTCAACCTGCGCATCGTGCACCTGGTG 157
|||||
QY 824 cggacccgagcgctgctgctcctccgagcgagcgagcgagcgagcgagcgagcgagc 883
|||||
Db 156 CGCGACCGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 97
|||||
QY 884 aacggcctgctgctgggacacacagcgcaagtgggtggagggcgagccctcagcctg 943
|||||
Db 96 AACGGCATGTGCTGGGCAACCAACGCGCAAGTGGGTGGAGGCGGACCTCACCCTGCGCTG 37
|||||
QY 944 attcgcaggtgtgctcgcagccacgctgcgcatcgc 978
|||||
Db 36 ATTCCGAGGTGTGCCCGACGACCGTGGCATCGC 2

RESULT 2
BE858652/c
LOCUS
DEFINITION
7401a09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305176 3'
similar to TR:Q9UB3 Q9UB3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ; mRNA sequence.
BE858652
ACCESSION

BE858652.1 GI:10373890
EST.
SOURCE
ORGANISM
Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 735)
NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGA), Tumor Gene Index
Unpublished (1998)
Contact: Robert Strausberg, Ph.D.
Email: cgapbsr@mail.nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/linL, send email to:
infoimage.llnl.gov
Seq primer: -400P from Gibco
High quality sequence stop: 425.
Location/Qualifiers
1. 735
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Brn23"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/note="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATATCTTTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."

BASE COUNT 135 a 221 c 253 g 120 t
ORIGIN

Query Match 30.9%; Score 631; DB 10; Length 735;
Best Local Similarity 93.5%; Pred. No. 1.2e-64;
Matches 677; Conservative 0; Mismatches 45; Indels 2; Gaps 2;

QY 300 cccggcagggccctcatcccccagcgccgagagatcgtgacagctgctgctgctg 359
|||||
Db 734 CCCGCNCAGGCGCCATCCACAGCAGNCGCGAGAACTCGTGTCACGTGGGTGCTGT 675
|||||
QY 360 cctcgtgctcgtgctcctctcttcttggccagctcttccagccagccacacgtct 419
|||||
Db 674 CTTCTGTGGCTCGGGCTCATCTCTNAGGGCCAGCTCTCAGCCAGCA-CGCCAGGCT 616
|||||
QY 420 tctacctgatggagcccgctggtgcatgtgtggaccacccctgtcgcagggcagcgcaa 479
|||||
Db 615 TCTACATGATGAAGCCCGCTGGCATGTGTGG-CCACCCTGTGCGAGGCGCAGCGGCAA 557
|||||
QY 480 cgcgcacatggcgctgcgacacgtgctcctctctctctctctctctctctctctct 539
|||||
Db 556 CGCTGCATGGCGCTGCGGACCTGATGCGCTCTATCTTTTGTGCGACATGGAGCTGT 497
|||||
QY 540 ttgatgcctacatgccacagagccgaacacgtctccgcctcttttccactggggcaagcgc 599
|||||
Db 496 TTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 437
|||||
QY 600 gcgcgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 659
|||||
Db 436 GCGCGCTGTGCTGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 377

QY 660 acgtatgaagacactgtgcaacgagcgagcgaattcagcctgcccgggagggcctgcgcgt 719
 Db 376 AGGTATGCAAGACACTGTGACGGCGCAGCCATTTCAGCTGCGCCGGAGGCTGCCGCT 317
 QY 720 cctacagcagcgtgtgtcctcaagaggtgcgtttcttcaacctgaggtgtctaccgc 779
 Db 316 CCTACAGCCAGCGTGTGCTCAGAGAGGTGCGCTCTTCAACCTGCAAGTGTCTACCGC 257
 QY 780 tgcctagcagcccgccgctcaacctgcgcatctgtgcacctgtgtgcgaccccgagcg 839
 Db 256 TGCCTCAGGACCGCCGCTCAACCTGCGCATGCTGTCACCTGCTGTCGCGACCCCGCGGCG 197
 QY 840 tgcctgcctccgggagcgcgccgagcgaactggtgcagcgacacgcatcgtgtgg 899
 Db 196 TGCCTGCGCTCCCGGAGCGCGCGCGGCGGATCTGGCAGCGGACACCGCATCGTGTGG 137
 QY 900 gcaccaacggaagtggtgagcgccgacccctcactgcctgctgattgcgaggtgtgcc 959
 Db 136 GCACCAAGGCAAGTGTGAGGCGCGACCTCACTCGCGCTGATTTGCGAGGTGTTC 77
 QY 960 gcagccagctgcgcatcgccgagcgccgacacactcaagccgaccccttctgcgcgcc 1019
 Db 76 GCAACCACTGCGGATTGCGGAAGCCACACACTCAACCGCCACCTTTTGGGGGGC 17
 QY 1020 gcta 1023
 Db 16 GNGA 13
 RESULT 3
 BF197521/c
 LOCUS
 DEFINITION
 7084408.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'
 similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 ACCESSION
 BF197521
 VERSION
 BF197521.1 GI:11086670
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 525)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/hcicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabsr@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 High quality sequence stop: 451.
 Location/Qualifiers
 1..525
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:3642903"
 /clone_lib="NCI_CGAP_Kid11"
 /lab_host="PH10B"
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
 Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo. "
 BASE COUNT 90 a 163 c 187 g 85 t
 ORIGIN
 Query Match 25.7%; Score 525; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2.7e-52;
 Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 443 catgtgtgaccacacctgtcgagggcgagcgcaacactgacatgcccgtgcgcgac 502
 Db 525 CATGTGTGACACACCTGTCTGCGAGGCGACGCGGCAACGCTACATGCGCGTGCAGC 466
 QY 503 ctgatgcgctctatctttttgtgcacatggacgtgtttgatgcctacatgcccagagc 562
 Db 465 CTGATGCGCTCTATCTTTTGTGCGACATGGACGTTTGTATGCTTACATGCCACAGAGC 406
 QY 563 gaaaacctgtccgcttttttaactgggcaacgagccgcgcgtgtgtgcgcgcgcgc 622
 Db 405 CAAAACCTGTCCGCGCTTTTCAACTGGGCACGAGCGCGCGCTGTGCTGCGCGCGCGC 346
 QY 623 tgcagcgctttcccccggagggcaccatcagcaagcagcgatgcaacacactgtgcaag 682
 Db 345 TGCAGCGCTTTCCTCCCGGAGGACCATATCAGCAAGAGGAGGATATGCAAGACACTGTGCAGC 286
 QY 683 cggcagccattcagcctggccgggagggcctgcctcctacagccagctgtgtgctcaag 742
 Db 285 CGGCAGCCATTTCAGCTGGCGCGGAGGCGCTGCCGCTCTACAGCCAGTGTGTCTCAG 226
 QY 743 gagtgcgctcttcaacctgcaggtgtcttaccgctgtcctcagcgccgcgcgcgc 802
 Db 225 GAGTGTGCGCTTCTTCAACTGCAGTGTCTACCGCGTGTCTACGCGAGCCCGCGCTCAAC 166
 QY 803 ctgcgcctgtgcacctgtgtgcgagcccgagcggtgtgcctcctcgggagggcgcg 862
 Db 165 CTGCGCATCTGTGCACCTGTGTGCGGCGACCGCGGCGGTGTGCTGCCGAGGCGCGC 106
 QY 863 ggcccgatactggcagcgacacggaatgtgtgtggcaccacacgcaagtgggtgag 922
 Db 105 GGCCCGATCTGCGCAGCGGACACACGGCATCTGTGTGGCACCACAGCGCAAGTGTGTGAG 46
 QY 923 gcgcacctcaactgcgcctgattgcgaggtgtgtgcgcgagccac 967
 Db 45 GCGACCCCTCACCTGCGCTGCTGATTCGCGAGGTGTGCGCGAGCGCCAC 1
 RESULT 4
 A1824100/c
 LOCUS
 DEFINITION
 A1824100 620 bp mRNA linear EST 21-DEC-1999
 wj46c01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405856 3'
 similar to TR:O75667 O75667 DJ71LL16.4 :contains PTR5.b2 PTR5
 repetitive element ;, mRNA sequence.
 ACCESSION
 A1824100
 VERSION
 A1824100.1 GI:5444771
 KEYWORDS
 EST.
 SOURCE
 human.
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 (bases 1 to 620)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/hcicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 JOURNAL
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Email: cgabsr@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
 info@image.llnl.gov
 High quality sequence stop: 451.
 Location/Qualifiers
 1..525
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:3642903"
 /clone_lib="NCI_CGAP_Kid11"
 /lab_host="PH10B"
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site: 1: Not I; Site 2: Eco RI;
 Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 917 Std Error: 0.00
Seq primer: -400P from gibco
High quality sequence stop: 490.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

1. 520
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="IMAGE:2405856"
/clone_lib="NCI_CGAP_Lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73b-Pac (Pharmacia) with a
modified polylinker: 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- oligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaldo."
195 G 227 g 89 t 3 others

```

BASE COUNT	106 a	195 c	227 g	89 t	3 others
ORIGIN					

Query Match	23.9%;	Score 487.6;	DB 9;	Length 620;
Best Local Similarity	87.4%;	Pred. NO. 5.5e-48;		
Matches 543:	Conservative	0;	Mismatches 77;	Indels 1;
				Gaps 1;

BEST LOCAL SIMILARITY 87.4%; FREQ NO: 3.3e 40;
Matches 543: Conservative 77: Indels 1: Gaps 1;

Qy 401 agccagcaccgcagcttcttcttactctgatggagcccgctgtgtggaccaccctg 460

Dh 620 AACGACACCCGACGGTCTCTTACNTAATGGACCCGGCTGGACGCTGTGGACCA -CTTG 562

QY 461 tgcgaggcagcgcggcaacgtgcacatggcgtgcggaacgtagcgcctatcttt 520
ph 561 ttgcgacgcacgcgcgcgaacgcgtgcacatggcgtgtcgcgcacgtgctgcgcctc 502

Qy 521 ttgtgcacatgacgtatttgatgccacagagccgaacctgtcggcttt 580
St 501 ccatcgcttcttcacgcccgggttatcccttacccttccccctccctc 442

Qy 581 ttcaactgggcaacgagccgcgctgtgctgccgcgcgcgccttccccga 640

QY 641 ggcaccatcagcaagcaggacgatatgcagagacactgtccacgcggcagccattcagcctg 700

Qy 701 gccgggggctgcgcgtcctacagcaagtgtgtcaaggagggtgcgccttcttaac 760

Qy 761 ctgcaggtgctctacccgcgtctcagcgaccccgcgctcaacctgcgcatctgcacctg 820

QY 821. gtgcgcacccgcgggcggcgtgctgcgcctcccgcggaggcggcggcccgatactaggcaagc 880

[illegible]

DD		141	GACAAAGGATCGTGCTGGGCACCAACGGACGAGGGTGGAGGCCCGACCCCGGCGCCTGGC	82
QY		941	ctgattcgcgaggtgtgccgcagccacgtgcgcatacgcgcaggcgccgcacactcaagcg	1000

db 81 GTGGTCCGGAGGTGTGCCGTAGCCACGTACGCATCGCAGAGGCCGCCACATCAAGCCG 22

Qy 1001 ccaccccttcctgcggcgccgc 1021
 ||||| |||||
 Db 21 ccaccccttntgcggcgccgc 1

RESULT	5
AW081348/C	

AW081348	436 bp	mrna	linear	EST 14-OCT-1999
LOCUS	xc41d06.x1	NCI_CGAP.C020	Homo sapiens	cDNA clone IMAGE:2586803 3'
DEFINITION	similar to TR:088199	088199	CHONDROITIN 6-SULFOTRANSFERASE. ;	MRNA
	sequence.			
ACCESSION	AWC81348			
VERSION	AWC81348.1	GI:6036500		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 436)			
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .			
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
JOURNAL	Unpublished (1997)			
COMMENT	Contact: Robert Strausberg, Ph.D.			

JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.

CONTACT: ROBERT STAUBERT, Ph.D.
Email: crapbs-r@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be found through the I M A G E Consortium/INL at:

www.bia.gov/your-voice-through-the-11-03-consolidated.html

Seq primer: -40UP from Gibco

High quality seq

	source
<hr/>	
FEATURES	

BASE	COUNT	source
64	136	g
57	169	g
		by Life Technologies."
		Normalized version of NCI_CGAP_Col8. Library constructed
		Normalized to Cot 500. Average insert size 1.1kb.
		Site.2: Notri; Cloned unidirectionally. Primer: Oligo dt.
		/note="Organ: colon; Vector: pCMV-SPORT6; Site.1: Sali;
		/lab_host="DH10B"
		/tissue_type="moderately differentiated adenocarcinoma"
		/clone_lib="NCI_CGAP_Col20"
		/clone="IMAGE:2586803"
		/db_xref="taxon:9606"
		/organism="Homo sapiens"

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

Query Match	20.6%;	Score 420.8;	DB 9;	Length 436;
Best Local Similarity	99.3%;	Pred. NO. 3.6e-40;		
Matches 433;	Conservative 0;	Mismatches 2;	Indels 1;	Gaps 1;

Qy 581 ttcaactgggaacagagccgcgctgtgctogcgcgccgcctgaagcgcttccccga 640
|||||
pB 436 TTCACTGGGCAACGAGCGCGCGCTGTACTCGCGGCCGCCTTCCCCGA 377

Qy 641 ggcaccatcagcaagcaggacgtatgcaagacactgtgcacggcggcagccattcagcctg 700
|||||
Db 376 ggcaccatcagcaagcaggacctatgcacagacactgtgcacggcggcagccattcagcctg 317

Qy 701 gccggaggcctgcgtcctacagccacgtagtgctcaaggaggtgcgtttcttcaac 760
|||||
|||

bB 216 ccccccccctcccccccttcacgccaccgtgatcgaagcgatgcgttcctccac 257
|||||

QY 761 ctgcagg-tgctctaacgcgctgtcagcagccccgcgcgtcaaacctgcgcattcgtgcacct 819

Qy	820	ggtgcgcagccgcgggcgtctgcgtcccgaggagcgcgccgatactgcacg	879
Db	196	GGTGGCGACCGCGGSCCGTGTGGCTCCCGGAGGGCGCGCCGATCTGGCAG	137
Qy	880	cgacaaacgcatcgtcgtggccacaaacggcaagtgggtgaggcgacccctacctgcg	939
Db	136	CGACACGGCATCGTGTGGCCACCAACGCAAGTGGTGGAGGCGCACCTCACTGGG	77
Qy	940	ctgtattcgcgagtggtgcgcgagccacatgcgcattcgcgagggcgcacactcaagcc	999
Db	76	CGTATTCGGAGGTGTGGCGGAGCCACCGATCGCGCATCGCGAGGGCGGCACACTCAAGCC	17
Qy	1000	gcacaccttctgcgc	1015
Db	16	GCCACCTTCTCTGGC	1

RESULT 6

AK009113	1923 bp	mRNA	linear	HTC 19-JAN-2002
LOCUS				
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:231003G18:carbohydrate (chondroitin 6/keratan sulfotransferase 4, full insert sequence).			
ACCESSION	AK009113			
VERSION	AK009113.1	GI:12843701		
KEYWORDS	HTC; CAP trapper.			
SOURCE	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:231003G18.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;			

REFERENCE

AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
KEYWORDS cDNA cloning

REFERENCES

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE

AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yanamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Todaya, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	Riken integrated sequence analysis (RISA) system-384-format
JOURNAL	sequencing pipeline with 384 multicapillary sequencer Genome Res. 10(11), 1757-1771 (2000)

REFERENCES

4 (sites)
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium

TITLE Functional annotation of a full-length mouse cDNA

JOURNAL OF
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGACGCTCTTTTCTTTTNNV 3'], cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 25.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGATGTCGAGTCTAATTAATAATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites 5' end: XhoI; 3' end: SstI. Host: SOLR

FEATURES

source

```

0.0253
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="MGD:MGI:1904558"
/db_xref="taxon:10090"
/clone="2310003G18"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cdna library"
/dev_stage="adult"
96.1262
/ene="chst4"
/ene="chst4"
/ene="chst4"
/note="carbohydrate (chondroitin 6/keratan)
sulfotransferase 4
data source:MGD, source key:MGI:1349479, evidence:ISS
putative"
/codon_start=1
/protein_id="BAB26078.1"
/db_xref="GI:12843702"
/translation="MMLLKGRLLMLGSGQVIVVALFTIMSVHRHLSQRESRPVMH
LVLSWSRSGSFVGLQFGHPDVFYLMPEAWHVMVTFTSTAKLHMARLLRSVFL
CDSMVFAYMNGPRKQSSLPQWQSRALCSAPCDFFPAHEISPKHKLCLCQOQP
DMVKAQRSGHGVFLKVKRFLSOLALYPLTDPSLNHVHVLPRPAVFSRRETHL
ELMDSHVILVGLSHLETKEEDQPYAMKIKTSQVDIVKAIQTLPEALQOQYILFV
DLVAPLAQATIRLYKFGVGLDFPLHQTIVNVTGKMGQGHAFTHARNALNVSQAWR
WSLPYKVSQLOACGAGMADLLGYLQVRSQOEGNLLDLLSSHHILQVFRFG"
445 a 533 c 492 q 453 t

```

**B
A
S
E C
O
U
N
T**

**O
R
I
G
I
N**

Query Match	17.2%;	Score 352.4;	DB 11;	Length 1923;
Best Local Similarity	50.8%;	Pred. No. 1.8e-32;		

Matches 613; Conservative 0; Mismatches 386; Indels 9; Gaps

QY 341 gtgcacgtgctgggtgctgtcctctgtggcgctggggtcattccttcttggggccagctcttc.400

Db 219 GTGCAATGGCTGGTGCTGTCTTCTGGCGGTTCAGGATCCCTCTTTTGTGGACAGCTTTTC 278

QY 401 agcagcaccgcgacgtttcttctacdtgatggagcccgccgtggcatgtgtggacaccctg 460

QY	714	gcgcgtctacagccacgtgggtgctcaaggagtgcgcttctcaacctgcaggtgctct	773
Db	426	GGACACCTTACAGCCACGTGGTGTGAATCGGTGGCTTCTTCAGCTGGAGTCCCTGT	367
QY	774	accgcgtctcaagcagaccgccgcgtcaacctgcgcatcgtgcacbtggtgcgacccgc	833
Db	366	ATCCGCTCTCGAGGACCCCAACCTTGAACCTCCGATCTCCACCTGGTCCGGGACCTC	307
QY	834	gggcgcgtcgtcgtcccgggaggcgcgggccgatactggcagcgacacaacgcacatg	893
Db	306	GGGCGGTGGCGGCTCCAGAGAACAGTCGCCAACGCCTGCTCCCGACGACGCCGTCA	247
QY	894	tgcctgggcacaaagcgcaagtgggtggaggcagccctcaactgcgctgattcgcgagg	953
Db	246	TCTTGGAGAACAGGCGCA---TGCGCGCGCGAGGTGCTGATCAGGTCAITGCAGAGA	190
QY	954	tgtgcgcagcaccgtgcgcactgcgcgagcgccacactcaaacgcgcacctctctgc	1013
Db	189	TC TGCGCGAGCCAGTGGCGATCAGCGAGAGGGCCACCGTGAAGCCCGCGCTTTCACA	130
QY	1014	gcgcgcgtaccgcctggtgcctctcggagacctggcgcgaggcgcttggcagagatcc	1073
Db	129	AAGCGCGTCAAAATGGTCCGCTTCGAGGACCTGGTCCGGGACCCGCTCGGGAGATCG	70
QY	1074	gcgcactctacgcttcaacgcgctgacctcaacgcacacgtcagagccttggtatccaca	1133
Db	69	CGGCATCTACAAGTCTGTCGTCTGGAGATGACCAAGCAGCTGGCGCAGTGGATTACC	10
QY	1134	acataacc	1141
Db	9	AGCTTTC	2

RESULT	8	
LOCUS	AG035205	
DEFINITION	Pan troglodytes DNA, clone: PTB-010M13.F, genomic survey sequence.	
ACCESSION	AG035205	
VERSION	AG035205.1 GI:16562078	
KEYWORDS	GSS; GSS (genome survey sequence).	
SOURCE	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male	
ORGANISM	BAC Library clone:PTB-010M13.F.	
	Pan troglodytes	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.	
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,	
	Totoki,Y., Watanabe,H. and Sakaki,Y.	
TITLE	BAC end sequences of Library PTB	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 657)	
AUTHORS	Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,	
	Totoki,Y., Watanabe,H. and Sakaki,Y.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical	
	and Chemical Research (RIKEN), Genomic Sciences Center (GSC);	
	1-7-22 Suehiro-cho,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan	
	(E-mail:chimbpgsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/;	
	Tel:81-45-503-9111, Fax:81-45-503-9170)	
	Clones are derived from the chimpanzee BAC library PTB This BAC end	
COMMENT	was generated during the R&D process and may have higher chance of	
	clone tracking errors.	

```

FEATURES
    source
        Sequencing: -21M13
        LIBRARY
            Vector      : pKSI45
            R.Site 1    : SacI
            R.Site 2    : SacI.
            Location/Qualifiers
                1..657
                    /organism="Pan troglodytes"
                    /db_xref="taxon:9598"

```

```

/clonename="PTB-010M13.F"
/sex="male"
/cell_type="lymphoblast"
/clonelib="PTB Chimpanzee Male BAC Library"
BASE COUNT      85 a   237 c   201 g   133 t   1 others
ORIGIN

Query Match          15.8%; Score 322.8; DB 12; Length 657;
Best Local Similarity 86.0%; Pred. No. 7.3e-29;
Matches 382; Conservative 0; Mismatches 57; Indels 5; Gaps 2

QY 201 gcccccagccgcgcgcatgtggctgccacggtttcccaagaacagcagtgaacagtctcc 260
Db 217 GCCCCCCAGCATTGACATGTGGCTGCCCGCGCTCCAGCACAGTGCACGGCGTCCT 276
QY 261 tcctggcacagaccctgccttcctgtcttcatcatcccccgccaggccctcatccc 320
Db 277 TCCTGGCGCAG---ACCCTCCCTCCCTCTTTCTGGTTTCCCGGCCAGGGCCCTCGTCCC 333
QY 321 cagccggcgaggagatcgttgccagctgctgtctgtcttccttgtggcctgggctcat 380
Db 334 CAGCAGGCGGGCAGCGCGCGCTGCATGTGCTGGTGCTCTCCCTGGTGCGCTCGGGCTCGT 393
QY 381 ccttcttggccagcgtcttccagccagcccgccagcgtcttctacctgatggagcccgct 440
Db 394 CCTTCGTGGGCCACTCTTCACACAGCACCCCGACGCTCTTCTACCTAATGGAGCCCGGT 453
QY 441 ggcatgtggaaccaacctgttcgaggcgagcgcgcaacgtgcacatggcctgctgcg 500
Db 454 GGCAGGTGTGACACACCCCTGTGCGAGGGCAGCGCGCAACGCTGCACATGGCTGTGCGGG 513
QY 501 acctgatcgctctattctttttgtgcaca--tggacgtgtttgatgctaataccaca 558
Db 514 ACCTGGTCGCTCTCGTCTCTCTGTGCGACAGTGGACAGTGTGTGATGCTATCTGCCTTG 573
QY 559 gagccgaacctgtccgccttttcaactgggaacagcgccgcgtgtgtctgcgcgcc 618
Db 574 CGCGCGGAACCTGTGCGACCTCTTCCATGTGGCCCTGAGCCGTGCACATGCTCGCCACC 633
QY 619 cgctgcagcgccctttccccagg 642
Db 634 CGCTGCAGTGCCTTCCCGAGG 657

RESULT 9
BF042384
LOCUS
DEFINITION
clone BP250022A10E10 5', mRNA sequence.
517 bp mRNA linear EST 10-OCT-2000
Bos taurus CDNA
```

RESULT	9	
BF042384		
LOCUS		
DEFINITION		517 bp mRNA linear EST 10-OCT-2000 BP250022A10E10 Soares normalized bovine placenta Bos taurus cDNA
ACCESSION	BF042384	
VERSION	BF042384.1	GI:10759439
KEYWORDS	EST.	
SOURCE	cow.	
ORGANISM	Bos taurus	
REFERENCE		Fukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos. 1 (bases 1 to 517)
AUTHORS	Lewin,H.A., Soares,M.B., Rebeiz,M., Pardini,J., Liu,L. and Larson ,J.H.	
TITLE	Bovine ESTs	
JOURNAL	Unpublished (2000)	
COMMENT	Contact: Lewin, H. A.	

W. M. Keck Center for Comparative and Functional Genomics
University of Illinois at Urbana-Champaign
340 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
61801, USA
Tel: 217 333 5998
Fax: 217 244 5617
Email: h-lewin@uiuc.edu

Funding for cattle EST sequencing was provided by the USDA National
Research Initiative, Animal Genome Resource Grant AG 99-3205-8534

to H. A. Lewin and J. E. Womack. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Triimi g: Cross_match from Washington University Genome Center PHRAP suite. This sequence is vector free and at least 200 bp in length.

PCR Primers

FORWARD: TAATACGACTCAGTATAGG
 BACKWARD: ATTAAGCCCTCAGTAAG
 Insert Length: 517 Std Error: 0.00
 Plate: BP250022A10 row: E column: 10
 Seq primer: ACGGATACCAATTCACAGGA
 High quality sequence stop: 517.

FEATURES

source
 1..517
 /location/Qualifiers
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="BP250022A10E10"
 /sex="female"
 /lab_host="DH10B"
 /note="Organ: placenta; Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI; The cDNA library was contributed by the Soares laboratory and it was constructed and normalized as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806."
 70 a 177 c 170 g 100 t

BASE COUNT
 ORIGIN

Query Match 15.1%; Score 308.4; DB 10; Length 517;
 Best Local Similarity 78.8%; Pred. No. 3.8e-27;
 Matches 410; Conservative 0; Mismatches 96; Indels 14; Gaps 3;
 QY 105 ttaagagacctgcttctcgtgaacacag-----atgcgagacatctccca 154
 Db 1 TTAAGAGACCTCTGTCATCCCTGGAGTGCAGGCCCTGAAAACCTCCAGGCCTCTCCCA 60
 QY 155 tggatttgatc-aggcttgacgtctctccagcagccctggagcgtggcccccagcgc 213
 Db 61 TGGACCTGTGATGAGGTGGGAGCCCTCCACAGCAGCCAGGAGTCTCGGGCGG 120
 QY 214 ccgcattggtgcacaggtttctccagcaagacagtgacagtgctctctctggcacagac 273
 Db 121 CGGCATGTGGCTCGCGCGGCTCTCCAGCAGCGGCTGACTGCCTCTCTGGCGCAG-- 178
 QY 274 caactgctctgctcttcatcatctccggccagggccctcctccacagccggcgga 333
 Db 179 -ACGGCCCTCTGCTCTCTCTGATCTCCGCGCCAGACTCCGCCCTCGCGCCAGCAGCA 237
 QY 334 ggatcgtgtgcacgtgctgtgctctcgtggtggtggtggtggtggtggtggtggtggt 393
 Db 238 GGAGCGGGTGCATGTGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 297
 QY 394 gctcttcagcagacacccagcgtcttctacctgatgagccgcgctggtggtggtggtggt 453
 Db 298 GCCTTCAGCCAGCACCCCGATGCTCTTCTACCTGATGAGACCCCGCTGGACGCTGCTGG 357
 QY 454 caccctgtcagggcagc 513
 Db 358 CGCCCTGTGCGAGGAGCGCCCTTGGCGCTGCACATGGCAGTGGCGACCTGGTGGCGTC 417
 QY 514 tatctttttgtgacatgacgtgtttgtgcttacctacatgccacagcgcgcgcgcgcgc 573
 Db 418 AGTCTTCTGTGGACATGATGTTGAGGGGCTACTGCGGTGGAGGGCGCAACCTGTC 477
 QY 574 cgctttttcaactgggcaagc 613
 Db 478 GGACCTCTTCATATGGCGGAGAGCGCGGCTCTGTGCTG 517

RESULT 10
 AII15260 568 bp mRNA linear EST 02-SEP-1998
 LOCUS AII15260
 DEFINITION ui43c07.y1 Sugano mouse embryo mewa Mus musculus cDNA clone

IMAGE:1885164 5' similar to SW:C6ST_CHICK Q92179 CHONDROITIN 6-SULFOTRANSFERASE ; mRNA sequence.
 AII15260
 AII15260.1 GI:3515584
 EST
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 568)
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 The WashU-HHMI Mouse EST Project
 Unpublished (1996)
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:969488
 Seq primer: custom primer used
 High quality sequence stop: 517.

FEATURES

source
 1..568
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /clone_lib="Sugano mouse embryo mewa"
 /dev_stage="embryo, 14 dpc"
 /lab_host="DH10B"
 /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTTGTG); Site_2: DraIII (CAGCAATGT); 1st strand cDNA was primed with an oligo(dT) primer [AATGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTTGTG, 3' site CAGCAATGT). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTCTGCTCTAAAGCTGGC and 3' end primer CGACCTGCAGCTCGAGCACA."
 80 a 189 c 173 g 126 t

Query Match 14.5%; Score 295.4; DB 9; Length 568;
 Best Local Similarity 75.4%; Pred. No. 1.2e-25;
 Matches 381; Conservative 0; Mismatches 121; Indels 3; Gaps 1;
 QY 202 gcccccagcccgcatgtgctgcccaggtttctccagcaagacagtgcagtgcctct 261
 Db 67 GGCCCTACCCAGAGATATGGCTACCCCGTTTCTCCAGCACATGTCTATCTGCTCTCT 126
 QY 262 cctggcagaccacctgctctcctgctcttctcctcctcctcctcctcctcctcctcct 321
 Db 127 GATGGTACAG--ACTGGCATCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTG 183
 QY 322 agcccgccgagcagatgctgacagtgctgctgctgctgctgctgctgctgctgctgct 381
 Db 184 AGCAGGCTTGGGAGCGGTGTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 243
 QY 382 ctctctggccagctcttcacgacacccacccacccacccacccacccacccacccaccc 441

```

Db 244 CTTCGTGGCCAGCTCTTACAGCAACACCCCGAGTCTTCTACCTGATGGAGCCGGCTTG 303
QY 442 gcatgtgtgacacacccctgtcgagggcagcgcaacgctgcacatgcctgtgcgcga 501
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 GCACGCTGGATACGTTTTCGAGGGCAGTGCCTCCGACACCTCCATGGCCGTGGTGA 363
QY 502 cctgatgcctctatcttttttggacatgacgtgttttgacctatgcctacatgcacag 561
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 CCGTATCCGCTCAGTCTTCTTACATGCACATGGACGATTTGATGCCCTACCTGCCCTGGCG 423
QY 562 ccgaacacctgcgcctcttttcaactggcgaacagcgcgcgcgcgcgcgcgcgcgcgc 621
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 CCGCAACATCTCGGATCTCTTCAGTGGCGGTGAGCCGCAITGTGCTACCTCCGCT 483
QY 622 ctgcgcgcctcttcccgaggcaccatcagcagcagcagcagcagcagcagcagcagcagc 681
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 CTGCGAAGCCTTCGCTCGTGGCAATCATCAGCAGCAGGAGGTGTGAAGCCTCTGTGGCGC 543
QY 682 gcggcagcattcagcctggcccg 706
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 544 AACCGCGCCCTTCGCGCTGGCTCAG 568

RESULT 11
LOCUS BI823850 954 bp mRNA linear EST 04-OCT-2001
DEFINITION 603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
ACCESSION BI823850
VERSION BI823850.1 GI:15935400
KEYWORDS EST..
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NTH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM1148 row: h column: 03
High quality sequence stop: 856.
Location/Qualifiers
1. 954
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5179826"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/notes="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site:1, NotI; Site:2, EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age range 23-27; 1
male lung, age 27; and 1 male testis, age 69. Library is
oligo-dr primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NIH_MGC Library."
BASE COUNT 199 a 297 c 250 g 208 t
ORIGIN

Query Match 13.9%; Score 284.2; DB 10; Length 954;
Best Local Similarity 73.9%; Pred. No. 1.9e-24;

```

```

Matches 388; Conservative 0; Mismatches 133; Indels 4; Gaps 2;
QY 335 gatcgtgtgacacacccctgtcgagggcagcgcaacgctgcacatgcctgtgcgcga 394
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 198 GAGCGGATGCACGTGTGGTTCCTCTTCCTGGCGCTCTGGCTCTCTTTTGTGGGCGAG 257
QY 395 ctcttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 454
CTTTTGTGGGCGAGCAGCCAGATGTTTCTTACCTGATGAGCCCGCTGGCAGTGTGATG 317
QY 455 accctgtgcagggcagcgcgcaacgctgcacatgcctgcgcgcgcgcgcgcgcgcgcgc 514
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 318 ACCITCAAGCAGACACCGGCTGGATGCTGCATGGCTGTGGGGATCTGATACGGGCC 377
QY 515 atcttttttggacatggacgtgttttgacctatgcacatgcacagag---ccgaacctg 571
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 GTCTTCTGTGGACATGAGCGTCTTTGATGCCCTACATGGAACCTGGTCCCCGGACAG 437
QY 572 tcgccttttcaactggcgaacgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 631
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 TCCAGCCTCTTTCAGTGGGAGAACACCGGCGCCTGTGTCTGCACCTGCTGTGACATC 497
QY 632 ttccccgaggcaccatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 691
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 498 ATCCACAGATGAATCATCCCGGGCTACTGCGAGCTCTGTGCGAGTCAACAGAGCC 557
QY 692 ttcagcctggccggaggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 751
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 TTTGAGGTGTGGAGAGGCGCTGCCCTCCTACAGCCACGCTGTGCTCAAGGAGGTGGCG 617
QY 752 ttcttcaactcagctgctctaccgcgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 811
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 618 TTCTTCAACCTGCAGTCCCTCTACCGGCTGCTGAAAGACCCCTCCCTCAACCTGCATATC 677
QY 812 gtgcacc-tgggtgcgcgacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 855
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 678 GTGCACCTGTGTCGGGACCCCGGCGCGTTCGTTCCCGAGA 722

RESULT 12
LOCUS CNS04QFN/c
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
129006 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL302540
VERSION AL302540.1 GI:8181872
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 849)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
Unpublished
2 (bases 1 to 849)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 849)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large

```


[illegible]

D	b	241	CAAGCCGACCCCTTCTGTGC	260
RESULT 14				
Bg107354				
LOCUS				
DEFINITION				
ACCESSTION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
FEATURES				
source				
<p>Bg107354 955 bp mRNA linear EST 30-JAN-2001 602290740F1 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:438588 5', mRNA sequence. Bg107354 Bg107354.1 GI:12601200 EST. human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 955) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999). Contact: Robert Strausberg, Ph.D. Email: cyapbs@remail.nih.gov Tissue Procurement: Louis Staudt, M.D., Ph.D. CDNA Library preparation by: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10067 row: o column: 09 High quality sequence stop: 634. Location/Qualifiers 1..955 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:438588" /clone_lib="NIH_MGC_85" /tissue_type="lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: lymph; Vector: pCMV-Sport6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 1.867 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC library." 196 a 292 c 277 g 190 t</p>				
BASE COUNT				
ORIGIN				
<p>Query Match 12.7%; Score 259.4; DB 10; Length 955; Best Local Similarity 77.7%; Pred. No. 1.4e-21; Matches 394; Conservative 0; Mismatches 96; Indels 17; Gaps 6;</p>				
Q	y	1170	aggcttcacataacttgcttagaatgcgcgaacgtctccaggctggcgcaagcgt	1229
D	b	30	AGCCTTCAGAGACTTGCTCGAGAAATGCGCTCAACGCTGCCAGGCTGGCGCATGCGC	89
Q	y	1230	tgccttcctaagaatccctgcgcgtcacaggagtgtgcgcggcgcgctgcagctctgg	1289
D	b	90	TGCCCTTTGCC--AAGATCCGCGCGGTGCAGGAAGTGTGCGCTGGTGGCTGCACGCTGTG	147
Q	y	1290	gtaccggcctgtgtactctgcgacacagcagcgtgacctcaacctggatctggtctac	1349
D	b	148	GCTACCGGCTGTGTACTCTGAGGACGACGCGG-CAACTCGCCCTTGATCTGGTCTGC	206
Q	y	1350	cacagggcccagaccatttcagctgggcatc--gcttgactgagaactctgggacctaga	1407
D	b	207	CACGAGGCTGAACGGCTTCACATTGGGCAATCACCCGCTCGCACCCCCGAAATAGT	266
Q	y	1408	gcaagcccc-gaaactgtgtgcgcagggcccggaagcagcagctgcatgtgtgaaaaggagc	1465
D	b	267	GGAGGCGCAGAGTGTAGACAGGGGCGTAGGCCCGGGAGAGAGTGCATGGTGCAGAGGGGCG	326
Q	y	1466	tggggcgcatgggaa-caggtccctactatcaacggagatttgggt-----cc	1515
D	b	327	TGGGCGCACGGAGAAGGTCCTATTAITGACCAAGGAGTTTTGTGGTACGACCTCCC	386
Q	y	1516	tcocctgaagtaagcaaggactgcacgtttctctctctctctctctctctctctctctt	1575
D	b	387	CCTCCCAAAGTAGCAAGGACTGCACGTTCTTCTCTCTCTGATCTTGTTTTCCCTTG	446
Q	y	1576	agtctctggagctgcctctctcatcaggtgcacgtctctcatgaaaaacaactcttgc	1635
D	b	447	AGTCTCTGGAGCTGCCCTTCTCATCAGTGCATCTTCAITGGAAGCAACTCTTGCCCC	506
Q	y	1636	tacotctctggcgaggagtaagt	1662
D	b	507	TGCTCTCTCTGGGCACAGGTGTGCGT	533
RESULT 15				
AZ405100/c				
LOCUS				
DEFINITION				
ACCESSTION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				
<p>AZ405100 433 bp DNA linear GSS 03-OCT-2000 1M0173108R Mouse 10kb plasmid UUGCLM library Mus musculus genomic clone UUGCLM0173108 R, DNA sequence. AZ405100 AZ405100.1 GI:10529113 GSS. house mouse. Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 433) Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly, .M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D.,Weiss,R. Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts Unpublished (2000) Contact: Robert B. Weiss University of Utah Genome Center University of Utah Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA Tel: 801 585 5606 Fax: 801 585 7177 Email: ddunn@genetics.utah.edu Insert Length: 10000 Std Error: 0.00 Plate: 0173 row: I column: 08 Seq primer: CACACAGGAACACAGCTATGACC Class: plasmid ends High quality sequence stop: 433. Location/Qualifiers 1..433 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="UUGCLM0173108" /clone_lib="Mouse 10kb plasmid UUGCLM library" /sex="Male" /lab_host="E. Coli strain XL10-Gold, Tl-resistant, F-" /note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptored DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (gi 4732114 gb AF129072.1), a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptored mouse DNA was annealed to</p>				

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on: September 24, 2002, 09:28:52 ; Search time 251.38 Seconds
          (without alignments)
          1997.275 Million cell updates/sec
```

```

Title: US-09-593-828-3
Perfect score:
Sequence: I ggctcaggtccactgtct.....aaaaaaaaaaaaaaaa 2044

```

Scoring table: IDENTITY NUIC

Gapop 10.0 ,

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

```

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 10%
                  Listing first 45

```

```
Database : Issued_patents_NA:*
1: /cgn2_6/ptodata/
2: /cgn2_6/ptodata/
3: /cgn2_6/ptodata/
4: /cgn2_6/ptodata/
5: /cgn2_6/ptodata/
6: /cgn2_6/ptodata/
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	400	19.6	2032	4	US-09-045-284A-1	Sequence 1, Appli
2	235.2	11.5	2156	2	US-08-899-514-1	Sequence 1, Appli
3	208.2	10.2	13865	2	US-09-009-217-11	Sequence 11, Appl
4	208.2	10.2	13865	3	US-09-009-656-11	Sequence 11, Appl
5	204.2	10.0	72928	3	US-09-009-913-1	Sequence 11, Appl
6	202.8	9.9	2713	2	US-08-916-901-6	Sequence 6, Appli
7	202.8	9.9	2713	4	US-09-154-602-6	Sequence 6, Appli
8	201.6	9.9	2115	1	US-08-395-800A-7	Sequence 7, Appli
9	201.2	9.8	22481	1	US-08-367-841A-43	Sequence 7, Appli
10	201.2	9.8	22481	5	PCM-US95-07201-43	Sequence 43, Appl
11	200.2	9.8	6669	3	US-09-213-971-5	Sequence 43, Appl
12	200.2	9.8	6669	3	US-08-800-949A-5	Sequence 5, Appli
13	200.2	9.8	6669	4	US-09-617-053A-5	Sequence 5, Appli
14	198.4	9.7	282	1	US-08-133-629-8	Sequence 5, Appli
15	198	9.7	17477	4	US-09-608-285A-42	Sequence 42, Appl
16	197.4	9.7	629	4	US-09-389-982-204	Sequence 204, Appl
17	196.8	9.6	2354	1	US-08-655-878-1	Sequence 1, Appli
18	196.8	9.6	20303	1	US-08-370-975B-6	Sequence 6, Appli
19	196.8	9.6	26764	1	US-08-370-975B-1	Sequence 1, Appli
20	196.6	9.6	12047	2	US-09-022-461-1	Sequence 1, Appli
21	196.4	9.6	1000	4	US-09-018-584A-32	Sequence 32, Appl
22	196.4	9.6	15977	4	US-09-608-285A-59	Sequence 59, Appl
23	194.1	9.5	8353	3	US-08-611-587-1	Sequence 59, Appl
24	193.6	9.5	631	4	US-09-385-982-354	Sequence 1, Appli
25	193.4	9.5	4823	2	US-08-457-224-5	Sequence 354, App
26	193.4	9.5	4823	2	US-08-484-237-20	Sequence 5, Appli
27	193.4	9.5	4823	3	US-08-999-927-5	Sequence 20, Appl
						Sequence 5, Appli

C 28	193.4	9.5	4823	4	US-08-461-919-5	Sequence 5, Appli
C 29	193.4	9.5	4823	5	PCT-US94-08806-28	Sequence 28, Appl
C 30	193.4	9.5	4823	5	PCT-US95-01829-5	Sequence 5, Appli
C 31	193.4	9.5	4823	5	PCT-US95-16020-5	Sequence 5, Appli
C 32	193.4	9.5	11288	4	US-08-646-361A-1	Sequence 1, Appli
C 33	193.4	9.5	11288	4	US-08-481-968A-4	Sequence 1, Appli
C 34	193.4	9.5	11288	4	US-08-154-712B-4	Sequence 4, Appli
C 35	193.2	9.5	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 36	193.2	9.5	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 37	193.2	9.5	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 38	192.6	9.4	246240	2	US-08-724-394A-20	Sequence 20, Appl
C 39	192.6	9.4	246240	2	US-08-724-394A-21	Sequence 21, Appl
C 40	192.6	9.4	246240	2	US-08-724-394A-22	Sequence 22, Appl
C 41	192.4	9.4	15977	4	US-09-608-285A-59	Sequence 59, Appl
C 42	191.6	9.4	36741	4	US-09-301-665-3	Sequence 3, Appli
C 43	191.4	9.4	1701	4	US-09-078-294-9	Sequence 9, Appli
C 44	191.2	9.4	28994	3	US-08-884-324-14	Sequence 14, Appl
C 45	191	9.3	14796	3	US-08-975-080-35	Sequence 35, Appl

ALIGNMENTS

```

RESULT      1
US-09-045-284A-1
; Sequence 1, Application US/09045284A
; Patent No. 6285192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-1

```

Query Match	19.6%;	Score 400;	DB 4;	Length 2032;
Best Local Similarity	63.6%;	Pred. No. 6.9e-69;		

[illegible]

```

Db      662  tttaggtgtggaagagcgtgccgtcctacagccacgtggtgtccaaggagtgccg  721
QY      752  ttcttcaactcagtgctctaccgcgtgtccagcagcccgcgctcaacctcgcatc  811
Db      722  ttcttcaactcagtgctctaccgcgtgtccagcagcccgcgctcaacctcgcatc  781
QY      812  gtgcacctggtgogagaccggcggtgctgctcccgaggagggcgggcccgata  871
Db      782  gtgcacctggtcgggagccccggcggtgttccgtgtcccgagagcgcacaaaggagat  841
QY      872  ctggcagcgcacaaagcgcatcgtctgggcacccaacgcgaagtgggtgagcgccacct  931
Db      842  ctcatgattgacgtcgcatgtgatggcgagcagcagcagaactcaagaaggagagac  901
QY      932  cactctgacctgattgcgaggtgtgcgcagcagccacgtgcgcacgtccgcagcgccacac  991
Db      902  caacctactatgtgatgcaggtcatctgtccaaagccagctggagatctacaagaccatc  961
QY      992  ctcaagcggccacctctcgtcgcgccgctaccgcctcgtgctgcgttcgagacctggcg  1051
Db      962  cagtccctggccaaagccctcgaggaacgtcactgtgtgtgctgtgagacctggct  1021
QY      1052  cgggagcgtgtgacagatccgcgcactctacgcttccacgacctgacctcaacgcca  1111
Db      1022  cgagccccgtggccagacttccgaaatgatgaattcgtggattggaaattcctggcc  1081
QY      1112  cagctcgagcgctggatccacaacatcccccaggggtcggggatcggaagccaatcgag  1171
Db      1082  catcttcagacctgggtgcataacatcacccgaggaagggcgatgggtga-----ccac  1135
QY      1172  gccttcatactcgtctaggaatgcgcgcacgtctctccagcgctgcgcgaecgcttg  1231
Db      1136  gctttccacaaatgccaggatgccettaatgtctccagcttgccgtggtcttg  1195
QY      1232  cccttcactaaagtctgcgtgcagggaggtgtgcgcgcgcgtcgagctgctgggc  1291
Db      1196  ccttatgaaggtttctgcacttcagaagcctgtggcgatgccatgaatttgctgggc  1255
QY      1292  taccgacctgtactctgcgcagcagcgtgacctcacctggatctgtg  1345
Db      1256  taccgcacgtcagatctgaacaagacagagaaacctgttctggtctcttg  1309

RESULT      2
US-08-899-514-1
; Sequence 1, Application US/08899514
; Patent No. 5910581
; GENERAL INFORMATION:
; APPLICANT: HABUCHI, OSAMI
; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
; TITLE OF INVENTION: FOR THE POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
; CITY: NEWPORT BEACH
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```


QY	947	cgcagaggtgtgcgcagccacgtgcgcattcgcgagggcgcacactcaagcgcacacc	1006
Db	1170	CGGGCAACTGGCGAGAG--CATCCGCTTCCGGGAGCTGGGGTTCGGCGAGCCGCC	1226
QY	1007	tctctgcgcgcgtaccgcctgtgccttcgaggaacctggcgcggagcctgtgcca	1066
Db	1227	TGCTTCGGGGCGCTTACATGCTGTGTGCTACGAGAGCTGCGACGCGCGCGCTGCAG	1286
QY	1067	gagatccgcgcactcaagccttcacggcctcagccctcagccacacagctcgaagcctgg	1126
Db	1287	AAGCGCCGGAGATGTACCGGTTCGCGGCATCCCCCTACCCCGCAGGTGGAAGACTGG	1346
QY	1127	atccacaacatcacccacgggtcgtgggatatggcaagccaatgagggcttcacatactcg	1186
Db	1347	ATCCAAAAGAACACAGCGCGCCACGACGGCA-----GGGGCATCTACTCCAG	1397
QY	1187	tetaggaaatgcgcgaagctctccagagcctggcgcacagcgttgcccttcaatgaatc	1246
Db	1398	CAGAAGAAGCTCTCGGAGCAGTTCAGAAAGTGGCGCTTCAGCATGCCCTTCAGGTGGCC	1457
QY	1247	ctgcgcgtgcgaagggtgtgcgcgcgcgcgtgcagctgctgggtacacggcctgtgtac	1306
Db	1458	CAGGTGTGTGCAGGCCCGCTGGCGCCCTGCCANGGCCCTTCGGGTACAACTGCGCGG	1517
QY	1307	tctcgggaccagcagcgtgaactcaactcgtgatctggtgctgcacgagggccacagccac	1366
Db	1518	GACGCGCGGCCCTCACCAACCGCTTCAGTCACGCTGTCTGGAGGAGGGGACCTCTGG	1577
QY	1367	ttcagctggyg	1376
Db	1578	GTCACTAGG	1587

RESULT 3
 US-09-009-217-11
 ; Sequence 11, Application US/09009217
 ; Patent No. 6132729
 ; GENERAL INFORMATION:
 ; APPLICANT: Thorpe, Philip E.
 ; APPLICANT: King, Steven W.
 ; APPLICANT: Gao, Boning
 ; TITLE OF INVENTION: COMBINED TISSUE FACTOR AND
 ; TITLE OF INVENTION: CHEMOTHERAPEUTIC METHODS AND COMPOSITIONS FOR COAGULATION
 ; TITLE OF INVENTION: AND TUMOR TREATMENT
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Arnold, White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/009,217
 ; FILING DATE: Concurrently Herewith
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/042,427
 ; FILING DATE: 27-MAR-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/036,205
 ; FILING DATE: 27-JAN-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/035,920
 ; FILING DATE: 22-JAN-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Hibler, David W.
 ; REGISTRATION NUMBER: 41,071

REFERENCE/DOCKET NUMBER:	UTSD:536
TELECOMMUNICATION INFORMATION:	
TELEPHONE: 512/418-3000	
TELEFAX: 512/474-7577	
INFORMATION FOR SEQ ID NO: 11:	
SEQUENCE CHARACTERISTICS:	
LENGTH: 13865 base pairs	
TYPE: nucleic acid	
STRANDEDNESS: single	
TOPOLOGY: linear	
US-09-009-217-11	
Query Match	10.2%; Score 208.2; DB 3; Length 13865;
Best Local Similarity	73.7%; Pred. No. I.I.e-31;
Matches 278; Conservative 0; Mismatches 98; Indels 1; Gaps	
QY 1669	taaatataaattgctgcaggcgcggtgcggtgcctcatgcctgtaatccagcat 1728
DB 8368	TTAAGAAAAATTATAGGAAAAAGAGCGTGCGCATGTTGCTGTGCTGTAAATCTAGAAC 8427
QY 1729	tttgagagcctgagcggtgatacaactgagtcagga-ttcaaacaccagocgtgccaa 1787
DB 8428	TTTGGGATGCTAAGCGGGTGGATACATTGAGGTCAGGAGTTCAAGACCAGCGCTGGCCAA 8487
QY 1788	catdgaacccctctctactaaaaatgcataaattagtcggtggtggtgcacatc 1847
DB 8488	CATGCTGAACCCCATCTCTACTAAAAATACAAAAATTAGCCAGCGCTGTTGTGGGTGC 8547
QY 1848	ctgtaatcccagctactttaggctgaggtggtgggaaatcaattggactccaaagtga 1907
DB 8548	CTGTAATCCAGCTACTTTGGGAGGCTGAGGAGGAGAAATCACTTCGACCAGGTGGAGCA 8607
QY 1908	ggttcagtaagctgaaatcatgcacatgcacctagcttgggtggcaaacgaaactct 1967
DB 8608	GGTTCAGTGAACTCAGATTGGCCCACTGCCTCCGGCTCGGTGCACAGAGCGAGACTCT 8667
QY 1968	atcaaaaaataattaaataattgttcaaaagtctgcgaaaaaataaaaaaataaaa 2027
DB 8668	GCTCTAAAAGAAAGGGAAAGAAAGAAAAAGAAAAAGAAAGAAAGAAAGGAAG 8727
QY 2028	aaaaaaaaaaaaaaaaaa 2044
DB 8728	CAGAGAAAGAAATATA 8744

```

RESULT# 4
US-09-009-656-11
; Sequence 11, Application US/09009656
; Patent No. 613730
; GENERAL INFORMATION:
; APPLICANT: Thorpe, Philip E.
; APPLICANT: King, Steven W.
; APPLICANT: Gao, Boming
; TITLE OF INVENTION: COMBINED TISSUE FACTOR AND FACTOR VIIa
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR COAGULATION AND TUMOR
; TITLE OF INVENTION: TREATMENT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/009,656
; FILING DATE: Concurrently Herewith

```



```

; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/916,901
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-916-901-6

```

```

Query Match          9.9%; Score 202.8; DB 2; Length 2713;
Best Local Similarity 80.6%; Pred. No. 8.7e-31;
Matches 274; Conservative 0; Mismatches 52; Indels 4; Gaps 3;

```

```

QY 1664 actgctaaataaataaattgtgtgcaggccgggtgcgtggtcgtatgctgtatccc 1723
Db 1086 ACTGTTTCATGAAAGAAAGATGGTGTCCGCCAGCGCGGTGCTCATGCTGTAATCCC 1145
QY 1724 agcattttgagaggtgagcggtggtggtacacctgaggtcagga-ttcaaaaccagcctg 1782
Db 1146 AGCACATTGGAGCGCGAGCGGGTGTGATCACCCTGAGGTGAGGAGTTCGAGACGAGCCTG 1205
QY 1783 gccacatagtgaaacccc-ctctctactaaaaatgc--aaaaattagtcggcggtgtg 1839
Db 1206 GCCAACATGTTGAAACCCCTGTCTACTAAAAATACAAAAAATTAGCTGGCGCGGTG 1265
QY 1840 gcacactcctgtaatccagctacttaggaggtgaagtgagggtgggaaatacaattggactcca 1899
Db 1266 GCSCATGCCCTGTAAATCCCACTACTCAGGAGGCTGAGGCGAGGAGAAATCAGTTGAACCCGG 1325
QY 1900 aaggtgaggttcagtaagctgaataatcagcactgcacccctgactgtggtggcaaacg 1959
Db 1326 GAGGCAGAGGTTGCAGTGAGCCGAGATCATGCCATTCGACATCCAGCCTGGTGACAGAGC 1385
QY 1960 aaaaactctataaaaaataaataaattgttcaaaa 1999
Db 1386 GAGACTCCATCTCAAAAAAATAAATAAATAAATATGAA 1425

```

```

RESULT 7
US-09-154-602-6
; Sequence 6, Application US/09154602
; Patent No. 6300472
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: RAB PROTEINS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,602
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/916,901
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0367 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2713 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-154-602-6

```

```

Query Match          9.9%; Score 202.8; DB 4; Length 2713;
Best Local Similarity 80.6%; Pred. No. 8.7e-31;
Matches 274; Conservative 0; Mismatches 62; Indels 4; Gaps 3;

```

```

QY 1664 actgctaaataaataaattgtgtgcaggccgggtgcgtggtcgtatgctgtatccc 1723
Db 1086 ACTGTTTCATGAAAGAAAGATGGTGTCCGCCAGCGCGGTGCTCATGCTGTAATCCC 1145
QY 1724 agcattttgagaggtgagcggtggtggtacacctgaggtcagga-ttcaaaaccagcctg 1782
Db 1146 AGCACATTGGAGCGCGAGCGGGTGTGATCACCCTGAGGTGAGGAGTTCGAGACGAGCCTG 1205
QY 1783 gccacatagtgaaacccc-ctctctactaaaaatgc--aaaaattagtcggcggtgtg 1839
Db 1206 GCCAACATGTTGAAACCCCTGTCTACTAAAAAATACAAAAAATTAGCTGGCGCGGTG 1265
QY 1840 gcacactcctgtaatccagctacttaggaggtgaagtgagggtgggaaatacaattggactcca 1899
Db 1266 GCSCATGCCCTGTAAATCCCACTACTCAGGAGGCTGAGGCGAGGAGAAATCAGTTGAACCCGG 1325
QY 1900 aaggtgaggttcagtaagctgaataatcagcactgcacccctgactgtggtggcaaacg 1959
Db 1326 GAGGCAGAGGTTGCAGTGAGCCGAGATCATGCCATTCGACATCCAGCCTGGTGACAGAGC 1385
QY 1960 aaaaactctataaaaaataaataaattgttcaaaa 1999
Db 1386 GAGACTCCATCTCAAAAAAATAAATAAATAAATATGAA 1425

```

```

RESULT 8
US-08-395-800A-7
; Sequence 7, Application US/08395800A
; Patent No. 5807732
; GENERAL INFORMATION:
; APPLICANT: LONE, JOHN B
; APPLICANT: LENNON, GREGORY
; APPLICANT: ROQUIER, SYLVIE
; APPLICANT: GIORGI, DOMINIQUE
; APPLICANT: KELLY, ROBERT J
; TITLE OF INVENTION: GDP-L-FUCOSE: BETA-D-GALACTOSIDE

```

;; TITLE OF INVENTION: 2-ALPHA-L-FUCOSYLTRANSFERASES, DNA SEQUENCES ENCODING THE
;; TITLE OF INVENTION: SAME, METHOD FOR PRODUCING THE SAME AND A METHOD OF
;; TITLE OF INVENTION: GENOTYPING A PERSON
;; NUMBER OF SEQUENCES: 22
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSPALT
;; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
;; CITY: ARLINGTON
;; STATE: VIRGINIA
;; COUNTRY: USA
;; ZIP: 22202
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/395,800A
;; FILING DATE: 28-FEB-1995
;; CLASSIFICATION: 435
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (703) 413-3000
;; TELEFAX: (703) 413-2220
;; TELEX: 248855 OPAT UR
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2115 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 64..1092
US-08-395-800A-7

Query Match 9.9%; Score 201.6; DB 1; Length 2115;
Best Local Similarity 69.7%; Pred. No. 1.4e-30;
Matches 287; Conservative 0; Mismatches 124; Indels 1; Gaps 1;

QY 1632 cccctaccttttggcgaggagtaagttaactgctaataataattaaattgtgcca 1691
DB 1187 CCATCTCTCTCTGTGAAGATCGTGGGCTGCAAGTAACAGAAATCTCAGTGACAGT 1246
QY 1692 ggcgggtgcggtgctcgtgtaatcccgagcatttgagagcgtgagcggtgga 1751
DB 1247 GGCTGGCGTGGTGGTCAIGCTGTANTGCTCGCACTTGGGAGGCCAGGCTGGTGA 1306
QY 1752 tcacctgaggtcagga-ttcaaaacagcctggccacatagtgaaacccctctctact 1810
DB 1307 TCACITGAGTCAAGGAGTTCAAGACTAGCTGGCCCAACATGGTGAACCCCATCTCGACT 1366
QY 1811 aaaaatgcaaaaattagtcggcggtggtggcacactcctgtaataccagctacttaggag 1870
DB 1367 AAAAATACAAAATTAAGCAGGCGGTGGTGACACTTGTAAATCCAGTACTCGGGAG 1426
QY 1871 gctgaggtgggaaatacaacttggaatccaaagtggaggttgagtaagctgaaatcatg 1930
DB 1427 GCTGAGGCACAGAGAATCACTTGACCCAGGAGCGGAGGTGAGTGCACAGATGGTG 1486
QY 1931 ccactgacccctagcttgggtggcaagcaaaactctatacaaaaaataataaatt 1990
DB 1487 CGCTGCACTCCAGCCTGGGTGACAGCAAGACTCCATCTCAAAAAAAGAAAAA 1546
QY 1991 tgttcaaaagtctgcgcaaaaaaataaaaaaataaaaaaataaaaaa 2042
DB 1547 AGAATGACAGGGTTCAAGACCAATAATCATGCATATCATATGACACAGAA 1598

RESULT 9
US-08-367-841A-43/C
; Sequence 43, Application US/08367841A

;; Patent No. 6319687
;; GENERAL INFORMATION:
;; APPLICANT: Chader, Gerald J.; Rodriguez,
;; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
;; APPLICANT: Tombran-Tink, Joyce
;; TITLE OF INVENTION: PIGMENT EPITHELIUM
;; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
;; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
;; NUMBER OF SEQUENCES: 43
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Morgan & Finnegan
;; STREET: 345 Park Avenue
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10154
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC Compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: WORDPERFECT 5.1
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/367,841A
;; FILING DATE: 30-DEC-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/257,963
;; FILING DATE: 07-JUN-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/952,796
;; FILING DATE: 24-SEP-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: DOROTHY R. AUTH
;; REGISTRATION NUMBER: 36434
;; REFERENCE/DOCKET NUMBER: 20264126US2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 758-4800
;; TELEFAX: (212) 751-6849
;; INFORMATION FOR SEQ ID NO: 43:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 22481 Base Pairs
;; TYPE: Nucleic Acid
;; STRANDEDNESS: Double
;; TOPOLOGY: Unknown
;; MOLECULE TYPE: Genomic DNA
;; FEATURE:
;; NAME/KEY: PI-147
;; LOCATION:
;; IDENTIFICATION METHOD:
;; OTHER INFORMATION: full length genomic
;; OTHER INFORMATION: sequence for PEDF plus flanking sequences.
US-08-367-841A-43

Query Match 9.8%; Score 201.2; DB 4; Length 22481;
Best Local Similarity 77.8%; Pred. No. 2.8e-30;
Matches 256; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 1660 agttaactgtataataataatgtgtgccaggccggtgcggtgcgtcactgtaa 1719
DB 9841 AGTGAAACATCCATCTCAAAAAAAGAGCCAGCGCAGTGGCTCAAACTGTAA 9782
QY 1720 tcccagcattttgagaggctgagcggtgtgataccctgaggtcaga-ttcaaaacag 1778
DB 9781 TCCAGCACCTTTCAGAGCGCGAGGTAGATCACCCTGAGGTTCAGAGCAT 9722
QY 1779 cctggccacatagtgaacccctctctactaaaaatgcaaaattagtcggtggt 1838
DB 9721 CCTGGCCCAACATGAGAAACCCCACTCTACTAAAAATACAAAATTAGCGGTGGT 9662
QY 1839 ggcacactcctgttaatcccgactacttaggagctgaggtgggaaatacacttgactcc 1898
DB 9661 GCGCATGCCTGTAATCCAGCTACTTTGGGAGGCTAAGCCAGGAGATCACTTGAATCCA 9602

Db 10118 GCTGAGGCAGGAGAATCGCTTGAACCTGGAAGGCAGAGGTTGCAGTGAGCCGAGATCTCA 10059

Qy 1931 ccactgcaccctagcttgggtggcaagcaaaactctatcaaaaaataa 1980

Db 10058 CCAGTGCACCTCCAGCCTGGGTGCAGAGCAAGACTCTGTCTCAAAAAAAA 10009

Search completed: September 24, 2002, 09:30:24
Job time: 12113 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 19:04:23 ; Search time 1398.08 Seconds
(without alignments)
1440.505 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtggctgcacgtttctc.....gctgggcacgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1173	100.0	2044	22	AA02699 Human glycosyl sul
2	1173	100.0	2170	22	AA02698 Human glycosyl sul
3	1173	100.0	160552	22	AA02697 Human glycosyl sul
4	1027.2	87.6	2988	21	AA02696 Human OREF1711
5	924.6	78.8	1647	24	AA02695 Human drug metabol
6	924.6	78.8	1694	22	AA02694 Human glycosyl sul
7	745.2	63.5	1937	24	AA02693 Murine intestinal-
8	745.2	63.5	1989	22	AA02692 Mouse glycosyl sul
9	400	34.1	1979	22	AA02691 Human full-length

10	400	34.1	2032	20	AA020792
11	398.4	34.0	1333	24	AA016947
12	377	32.1	2065	21	AA024211
13	354	30.2	1926	20	AA020793
14	297.6	25.4	877	22	AA01803
15	297.6	25.4	877	22	AA01803
16	235.2	20.1	2156	19	AA03921
17	196.8	16.8	2354	18	AA021200
18	153.8	13.1	3029	22	AA017922
19	148.8	12.7	2393	24	AA01699
20	148.8	12.7	2409	20	AA017821
21	139	11.8	1458	19	AA036418
22	133	11.3	2150	20	AA018043
23	64.6	5.5	43280	18	AA018043
24	57.8	4.9	731	20	AA024563
25	57.8	4.9	731	21	AA018043
26	57.8	4.9	1644	23	AA054247
27	57.2	4.9	30001	18	AA0161016
28	57.2	4.9	30001	20	AA0161016
29	56.8	4.8	303	24	AA0161016
30	56.8	4.8	303	24	AA0161016
31	55	4.7	114955	20	AA053491
32	54.8	4.7	58857	21	AA053491
33	54.6	4.7	4689	21	AA017299
34	54.6	4.7	36778	21	AA017318
35	54.6	4.7	37948	21	AA017285
36	54.6	4.7	38506	21	AA017533
37	54.6	4.7	38506	21	AA017533
38	54	4.6	795	19	AA055830
39	54	4.6	109519	22	AA055830
40	53.8	4.6	1925	20	AA008693
41	53.4	4.6	53789	19	AA009024
42	53.4	4.6	65140	22	AA021187
43	53.4	4.6	125401	22	AA017184
44	53.2	4.5	1683	22	AA017186
45	52.8	4.5	1926	21	AA003820
					AA005254

ALIGNMENTS

RESULT 1

AA02699 standard; cDNA; 2044 BP.

XX AA02699;

DT 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocarditis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1; ss.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 218..1390

FT /tag= a

FT /product= "Human glycosyl sulfotransferase-4alpha

(GST-4alpha)"

FT /note= "CDS is specifically claimed as SEQ ID NO: 4

XX in claim 6 (page no: 41) of the specification"

PN WO200106015-A1.

ID	AA02697 standard; DNA; 160552 BP.
XX	AA02697;
XX	02-MAY-2001 (first entry)
XX	Human glycosyl sulfotransferase-4 (GST-4) genomic DNA.
DE	
XX	Human; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
KW	therapy; selectin binding inhibitor; gene therapy; inflammation;
KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW	polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW	chromosome 16q23.1; ds.
XX	
XX	Homo sapiens.
XX	
FH	Key Location/Qualifiers
FT	exon 32847..32922
FT	/tag= a
FT	/number= 1
FT	/label= 4a_5U4
FT	32923..35592
FT	/tag= b
FT	/cons_splice= (5'site:NO, 3'site:YES)
FT	35593..35674
FT	/tag= c
FT	/number= 2
FT	/label= 4a_5U3
FT	35675..45093
FT	/tag= d
FT	45094..45185
FT	/tag= e
FT	/number= 3
FT	/label= 4a_5U2
FT	45186..46633
FT	/tag= f
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	46634..46700
FT	/tag= g
FT	/number= 4
FT	/label= 4a_5U1
FT	46701..47938
FT	/tag= h
FT	/cons_splice= (5'site:YES, 3'site:NO)
FT	47939..47946
FT	/tag= i
FT	/number= 5
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	47939..47955
FT	/tag= j
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	47956..49128
FT	/tag= k
FT	/product= "Human glycosyl transferase-4alpha
FT	(GST-4alpha)"
FT	49129..49746
FT	/tag= l
FT	83257..83347
FT	/tag= m
FT	/label= 4a_5U2
FT	83348..96412
FT	/tag= n
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	96413..96484
FT	/tag= o
FT	/label= 4a_5U1
FT	96485..98456
FT	
FT	intron

FT	/tag= p
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	98457..99968
FT	/tag= q
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	98457..98473
FT	/tag= r
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	98474..99661
FT	/tag= s
FT	/product= "Human glycosyl transferase-4beta
FT	(GST-4beta)"
FT	99662..99968
FT	/tag= t
WT	WO200106015-A1.
PN	25-JAN-2001.
XX	19-JUL-2000; 2000WO-US19741.
XX	20-JUL-1999; 9PUS-0144594.
PR	13-JUL-2000; 2000JUS-0593828.
XX	(REGC) UNIV CALIFORNIA.
XX	Rosen SD, Lee JK, Hemmerich S;
PI	WPT: 2001-138471/14.
DR	P-PSDB; AAY72639, AAY72640.
XX	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT	diagnostic and therapeutic agent screening applications -
XX	Example 1; Page 62-104; 128pp; English.
PS	The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
CC	DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
CC	chromosome 16q23.1.
CC	GST is a type 2 membrane protein useful for inhibiting a binding event
CC	between a selectin and a selectin ligand, which comprises contacting the
CC	selectin with a non-sulphated selectin ligand, GST and a small molecular
CC	agent that inhibits the sulphation activity of GST. GST is also useful
CC	in inhibiting a selectin mediated binding event. GST is useful in gene
CC	therapy to treat disorders such as acute or chronic inflammation,
CC	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC	nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC	glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC	disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC	anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC	dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC	syndrome, infantile eczema, psoriasis, lichen planus, allergic rhinitis,
CC	bronchial asthma, hypersensitivity, rheumatic fever, and tissue rejection
CC	during transplantation.
XX	Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;
SQ	

Query Match	100.0%; Score 1173; DB 22; Length 160552;
Best Local Similarity	100.0%; Pred. No. 5.5e-217;
Matches 1173; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1 atgtggctgccacggttctccagcaagcagtgacagtgcctctcctctggcacagaccac 60
Db	47956 atgtggctgccacggttctccagcaagcagtgacagtgcctctcctctggcacagaccac 48015
QY	61 tgcctcctgcttttcattcatctcccggcccgacctatccccagcgcgcgcaggat 120
Db	48016 tgcctcctgcttttcattcatctcccggcccgacctatccccagcgcgcgcaggat 48075
QY	121 cggtgtgcacgtgcctggtgctctcctcgttgcgctcggttcctcttcttggccagctc 180

Query Match	87.6%	Score 1027.2	DB 21	Length 2988
Best Local Similarity	92.7%	Pred. NO. 4.9e-189		
Matches 1090	Conservative	0	Mismatches 83	Indels 3
Gaps				
QY 1	atgtggctgcacagttctccagcaagacagtgacagtgctctctctggcagacacacc	60		
Db				
1159	atgtggctgcacagttctccagcaagacagtgacagtgctctctctggcagacacacc	1218		
QY 61	tgcctctgtcttctaatctcccgccagggccctctatcccaagccggcgaggat	120		
Db				
1219	tgcctctgtcttctaatctcccgccagggccctctatcccaagccggcgaggat	1278		
QY 121	cgatgcacgtgtgggtctctctgtggcgctgggctcatctcttctgggcagctc	180		
Db				
1279	cgatgcacgtgtgggtctctctgtggcgctgggctcatctcttctgggcagctc	1338		
QY 181	ttcagccagaccccgagcttctacctgatgagcccgctggcatctgttggaacacc	240		
Db				
1339	ttcagccagaccccgagcttctacctgatgagcccgctggcatctgttggaacacc	1398		
QY 241	ctgtgcagggcagcgcgcaacgtgcacatggcgtgcgcagacctgatcgctctac	300		
Db				
1399	ctgtgcagggcagcgcgcaacgtgcacatggcgtgcgcagacctgatcgctctac	1458		
QY 301	tttttgtcgacatgacgtgtttgatgcctacatgccacag--ccgaaacctgtcc	357		
Db				
1459	tttttgtcgacatgacgtgtttgatgcctacatggaaacctgtcccgcgagacgtcc	1518		
QY 358	goccttttcaactgggaacgagcgcgctgtactgcgcgcgcgccttgacgacctt	417		
Db				
1519	agcctcttcagtggagaaacgcgcgcgcgcctgttctgcacctgcctgtacatcac	1578		
QY 418	ccccgaggaacctcagcaagcaggacgttatcaagacactgtgcacgcggcagccatc	477		
Db				
1579	ccaaagatgaatacatccccgggtcactgcaggctctgtgcagtcaacagcccttt	1638		
QY 478	agcctggcccgggagccctgcgcctctctacagccagctggtgtctcaagaggtgcctc	537		
Db				
1639	gaggtgggtggaagagcctgcgcctctctacagccagctggtgtctcaagaggtgcctc	1698		
QY 538	ttaacctgcaggtgctctaacctgctgcacgcacccgcgcctcaacctgcgcactcgt	597		
Db				
1699	ttaacctgcaggtgctctaacctgctgcacgcacccgcgcctcaacctgcgcactcgt	1758		
QY 598	cacctggtgcgcacccgggocgtgtgcctcccgaggcgcgcccgactactg	657		
Db				
1759	cacctggtgcgcaccccgggcgtgtgcctcccgaggcgcgcccgactactg	1818		
QY 658	gcacgcacaacgcacatgctgtggcaccacacgcgcaagtgggtggagccgacctcac	717		
Db				
1819	gcacgcacaacgcacatgctgtggcaccacacgcgcaagtgggtggagccgacctcac	1878		
QY 718	ctgcgcctgattgcgaggtgtgccgcagccactgcgcactgcgcagggcgccacactc	777		
Db				
1879	ctgcgcctgattgcgaggtgtgccgcagccactgcgcactgcgcagggcgccacactc	1938		
QY 778	aaagcccaaccttctctgcggccgactaacgacctggtgcgttcaaggaactggcgcg	837		
Db				
1939	aaagcccaaccttctctgcggccgactaacgacctggtgcgttcaaggaactggcgcg	1998		
QY 838	gagccgtggcagatccgcacactctacgctctcacgcgctgcacctcaagccacag	897		
Db				
1999	gagccgtggcagatccgcacactctacgctctcacgcgctgcacctcaagccacag	2058		
QY 898	ctcagggcttggtatccacaacatcaacccaggttcggggaatcggaagccaatcgagcc	957		
Db				
2059	ctcagggcttggtatccacaacatcaacccaggttcggggaatcggaagccaatcgagcc	2118		
QY 958	ttccatactctgtatggaatgcgcacacgtctccagacctggcgccacggttgcct	1017		
Db				
2119	ttccatactctgtatggaatgcgcacacgtctccagacctggcgccacggttgcct	2178		

useful for diagnosing, treating, or preventing disorders associated with aberrant expression of DME such as allergy, anemia, asthma, infertility -

Claim 5; Page 139; 143pp; English.

The invention relates to human drug metabolising enzymes referred as DME and nucleic acid molecules encoding such enzymes. Polynucleotides of the invention are useful for assessing toxicity of test compounds and in gene therapy. Sequences of the invention are useful in the diagnosis, prevention and treatment of autoimmune/inflammatory disorders such as acquired immune deficiency syndrome (AIDS), adult respiratory distress syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis, uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic infections; cell proliferative disorders such as actinic keratosis, arteriosclerosis, atherosclerosis, Buchanne and Becker dystrophy, hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as Addison's disease, adenomas, infarction associated with pregnancy, aneurysms, vascular malformations; eye disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa; metabolic disorders such as Addison's disease, cystic fibrosis, diabetes, gout, glycogen storage diseases, hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies, Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis, hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental disorders. The present sequence is human DME-5 protein cDNA.

Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

Query Match 78.8%; Score 924.6; DB 24; Length 1647;
Best Local Similarity 87.8%;
Pred. NO. 2.7e-169;
Matches 1021; Conservative 0; Mismatches 139; Indels 3; Gaps 1;

1 atgtggtgcacaggtttctccagcaagaacagtgcagtgctctcctggcacagacc 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 atgttgcgtgcgcgtctccagcacagcagtcgacctctcctgtgcgcag---acc 57
QY tgcctctgtcttcatcatctcccgcacaggccctcatccccagcgcggcgaggat 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
58 ttccctcctcttcttggtttcccgccaggccctcgctcccacagcaggcgcgagcg 117
QY 121 cgtgtgcactgtgtgtgttoctgtggtcgtgcgcgtgcattcttctgggccagctc 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
118 cgcgtgcactgtgtgtgtctcgtggtcgtgcgcgtgcgtcgtcctcgtggccaactc 177
QY 181 ttcagccagcaccocccagctctctactcgtatgagaccgcggtggcatgtgtgacacc 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
178 ttcaaccagcaccocgaagctcttctaactaatggagccgcgtggcactgtggacacc 237
QY 241 ctgtccaggcgacgvcggcaacgctgcacatggcgtgcgcacactgtagcgtctatc 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
238 ctgtccaggcgacgcgcgcgaacgctgcacatggctgtgcgcacctggfgcctcogtc 297
QY 301 ttttttgcacatgacgctgtttgatgcctacatgccacagagccgaacacctgtccgc 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
298 ttctgtgcacaatggacgtgttgtagcctatactccttggcgcgcgaacctgtccgac 357
QY 361 ttttttaactgggcaacagcagccgcgcgtgtgtcgcgcgcgcctgcagcgccttccc 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
358 ctcttcagttgggcgtgagccgtgcactgtgtcgcaccaaccgctgcagtgcctttccc 417
QY 421 cgagtcaccatcagcaagcaggaagctatgcaagaacactgtgcacgcggcagccattcagc 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
418 cgagcgccatcagcagcagcgcgtgtgcaagccactgtgcgcgcggcagctccttcacc 477

QY	481	ctggcccgaggcctgcggtccttacagccaagtgggtgtcaaggagggtgcgttcttc	540
Dd	478	ctggcccgaggcctgcggtccttacagccaagtgggtgtcaaggagggtgcgttcttc	537
QY	541	aacctgcaggtgctatcccgtgtctcagcgacccccgcgtcaacctgcgcatactcgtgcc	600
Dd	538	aacctgcaggtgctatcccgtgtctcagcgacccccgcgtcaacctgcgcatactcgtgcc	597
QY	601	ctggtgcgcaccgcggcgtgctgcgtctccgggagcgcgccgcgcatactcgtgcc	660
Dd	598	ctggtgcgcaccgcggcgtgctgcgtctccgggagcgagcagcgaaggtcgtggcg	657
QY	661	cgcacaacgcgatctgtctgggcaccaaacgcgaagtgggtgagcgccaccctcacctgc	720
Dd	658	cgtgacaacgcgatcgtctgggcaccaaacgcgaagtgggtgagcgccaccctcacctgc	717
QY	721	cgcttgattcgagaagtgtccgcagccaagtgcgcatacgcgcagccocacactcaag	780
Dd	718	cgcttggtgcgcagagtgtgcgtgagccaagtacgcatacgcgcagccocacactcaag	777
QY	781	cgcgcacacttctcgcgcgcgcatacgcgcgtggtgcgttcagaaactgcgcggag	840
Dd	778	cgcgcacacttctcgcgcgcgcatacgcgcgtggtgcgttcagaaactgcgcggag	837
QY	841	ccgtgtgcagagatccgcgcacttacgccttcacggcgtcaccctcacgcacagctc	900
Dd	838	ccgtgtgcagaaaatcgtgcgtctacgccttcactgggtcagttcaagccacagctc	897
QY	901	gagcctggatccaacaatacccccaggtctggggatcggcaagccaatcagagccttc	960
Dd	898	gagcctggatccaacaatacccccaggtcctggacctgtggacctgtgcgcgcgaagccttc	957
QY	961	catacttcgttaggaatgcgcgaacgtctcccagcctgcgcgcgcgtgcgcacagcttcgccttc	1020
Dd	958	aagacttcgtccagaatgcgtcaacgtctcccagcctgcgcgcgcgtgcgcacagcttcgccttc	1017
QY	1021	actaagatcctgcgcgtgcaggaggtgtgcgcgcgcgcgtgcgcacgtcgtgggttaacgcg	1080
Dd	1018	gccaaagtccgcgcgtgcaggaaactgcgcgtgtgcgtgcagctcgtgggttaacgcg	1077
QY	1081	cctgtgtactctgcgcacagcagcgtgaacctcacctcaccctggatctggtgtccacaggc	1140
Dd	1078	cctgtgtactctgcgcacagcagcagcgaacctcgccttgatctggtgtccacaggc	1137
QY	1141	ccagaccacttcagctgggcac	1163
Dd	1138	ctgaacggttcacttgggcac	1160
RESULT	5		
AAD02700			
ID	AAD02700 standard; cDNA; 1694 bp.		
XX AC	AAD02700;		
XX AC			
DT	02-MAY-2001 (first entry)		
XX DE	Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.		
XX DE			
KW KW	Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation;		
KW KW	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;		
KW KW	polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;		
KW KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;		
KW KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;		
KW KW	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;		
KW KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;		
XX XX	asthma; hypersensitivity; rheumatic fever; tissue rejection;		
XX XX	chromosome 16q23.1; ss.		
OS	Homo sapiens.		

PH	Key	Location/Qualifiers	
FT	5'UTR	9..188	
FT		/*tag= a	
FT	CDS	189..1376	
FT		/*tag= b	
FT		/product= "Human glycosyl sulfotransferase-4beta (GST-4beta)"	
FT		/note= "CDS is referred as SEQ ID NO:21 in brief description of the figures (page no: 4)."	
FT		1377..1694	
FT	3'UTR	/*tag= c	
FT			
XX	WO200106015-A1.		
PN	25-JAN-2001.		
XX	19-JUL-2000; 2000WO-US19741.		
XX	20-JUL-1999; 99US-0144694.		
PR	13-JUL-2000; 2000US-0593828.		
XX	(REGC) UNIV CALIFORNIA.		
PA	Rosen SD, Lee JK, Hemmerich S;		
XX	WPI; 2001-138471/14.		
DR	P-PSDB; AA72640.		
XX	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications		
PT	Claim 6; Fig 4A; 128pp; English.		
XX	The present sequence is human glycosyl sulfotransferase-4beta (GST-4 beta) cDNA. GST-4 gene is found on chromosome 16q23.1.		
CC	GST is a type 2 membrane protein useful for inhibiting a binding event between a selectin and a selectin ligand, which comprises contacting the selectin with a non-sulphated selectin ligand, GST and a small molecular agent that inhibits the sulphation activity of GST. GST is also useful in inhibiting a selectin mediated binding event. GST is useful in gene therapy to treat disorders such as acute or chronic inflammation, systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes, glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious anaemia, demyelinating diseases, cirrhosis, ulcerative colitis, dermatitis, myocarditis, regional enteritis, adult respiratory distress syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis, bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection during transplantation.		
CC	Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other:		
XX			
XX	Query Match	78.8%; Score 924.6; DB 22; Length 1694;	
XX	Best Local Similarity	87.8%; Pred. No. 2.7e-169;	
XX	Matches 1021; Conservative	0; Mismatches 139; Indels 3; Gaps 1;	
QY	1	atgtgctgcacaggttctccagacagatgacagtgcctctctctctggcacagaccac 60	
Db	189	atgtgctgcacaggttctccagacagatgacagtgcctctctctctggcgagcag---acc 245	
QY	61	tgcctctgtcttctcatctccgcagcgccctccatcccccagcgccgagagat 120	
Db	246	tctctctctcttctgttctccgcagcgccctccatcccccagcgccgagagc 305	
QY	121	cgtgtgcacgtgtgtgtctctctgttgcgtctggcgtctcattcttcttggcgagctc 180	
Db	306	cgcgtgcagtgcgtgtgtctctctgttgcgtctggcgtctggcgtctctcttcttggcgagctc 365	
QY	181	ttaagcagacacccgacgttcttctacgtgatgagccggtgacatgtgtgacaccc 240	
Db	366	ttaacacagaccccgacgttcttctacgtgatgagccggtgacatgtgtgacaccc 425	

RESULT 7
 AAS16948
 ID AAS16948 standard; cDNA; 1937 BP.
 XX
 AC AAS16948;
 XX
 DT 12-MAR-2002 (first entry)

[illegible]

(INCY-) INCYTE PHARM INC.

Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
Hillman JL, Azimzai Y;
WPI: 2000-256996/22.
P-PSDB; AAY79219.

Human transferase proteins useful for preventing, diagnosing and treating cancers and developmental, gastrointestinal, genetic, immunological, neurological, reproductive and smooth muscle disorders -

Claim 9: Page 104-105; 113pp; English.

The present sequence is that of cDNA clone 2617407CB1 encoding human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human transferase proteins of the invention (see AAY79209-23). The clone was isolated from gall bladder cDNA library GB1ANOR01. TRNSFS-11 is expressed in dermatologic and gastrointestinal tissues, especially those associated with inflammation and cell proliferation. It shows homology to mouse N-acetylglucosamine 6-O-sulfotransferase. A polynucleotide comprising nucleotides 264-323 or 1279-1331 of the present sequence can be used as a DNA probe. The new human transferases and polynucleotides can be used in the diagnosis, prevention and treatment (including gene therapy and antisense therapy) of cancer, developmental disorders, gastrointestinal disorders, genetic disorders, immunological disorders, neurological disorders, reproductive disorders, and smooth muscle disorders.

Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other:

Query Match	32.1.8;	Score 377;	DB 21;	Length 2065;
Best Local Similarity	63.4%;	Pred. No. 7.3e-64;		
Matches 644;	Conservative 0;	Mismatches 360;	Indels 11;	Gaps 4
QY	118	gacgtgtgcaactgtgtgtgtctctgttgagcgtcggctctcaactcttcttgggcag	177	
Db				
Db	294	gagcgatgcagtggtgtctctctgagcgtctgctctcttcttgggcag	353	
QY	178	ctcttcagccagcaccccgacgtctcttacctgatggagccgcgtggcatgtgtggacc	237	
Db				
Db	354	ctttttggcgagcacccagatgtttctacctgatggagccgcgtggcaggtgtggatg	413	
QY	238	acctgtcggaggcagcggcggaacgtgtcacatggcgtgcggacctgatgcgtct	297	
Db				
Db	414	acctcaagcagcagcccgctggatgtctgcacatggctgtcgggtatctgatacgggcc	473	
QY	298	atctttttgagacatgagcgtgtttgatgctacatgcacagag--ccgaaacctg	354	
Db				
Db	474	gtctcttctgcacatgagctgtttgatgctctacatggaaacctgtgtcccgagacag	533	
QY	355	tccgccttttttaactgggcaacgagcgcgcgtgtctgcgcgcgcgcctgcagccc	414	
Db				
Db	534	tccagccttttcaagtgggagaacgcggccctgtgtctgcactgctgtgcacatc	593	
QY	415	tttcccgaggcaaccatacgaagcagcagtatgcacagacactgtgcacgcggcag--cc	473	
Db				
Db	594	-atcccacagatgaatcatcccccgggtccactgcagctctgtgcagtcacacgcc	652	
QY	474	atttcagctggccgggagcctgcgcgtctcacagccactgtgtgtctcaaggaggtgcg	533	
Db				
Db	653	tttgaagtgtgtggagaagcctgcgcgtctctcacagccacgtgtgtccagggaggtgcg	712	
QY	534	ctttctcaacctgcaggtgtctacccgtgtctcagcagccgcgcctcaacctgcgcgt	593	
Db				
Db	713	ctttctcaacctgcagtcctctacccgtgtgtgaagacccctccctcaacctgcata	772	
QY	594	cgtgcacctgtgtgcgcacccgcgggcgcgtgtgtgcgtctcccgaggcgcggcccgat	653	
Db				
Db	773	cgtgcacctgtgtccggagccccgcggcgtgtgtccgtgtcccgagaacgcacaaaggaga	832	

[illegible]

RESULT 13

AAZ20793	
ID	AAZ20793 standard; DNA; 1926 BP.
XX	
AC	AAZ20793;
XX	
DT	08-DEC-1999 (first entry)
XX	
DE	Mouse glycosyl sulfotransferase-3 coding sequence.
XX	
KW	Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing
KW	selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW	secondary lymph organ; ss.
XX	
OS	Mus sp.
XX	
PN	W09949018-A1.
XX	
PD	30-SEP-1999.
XX	
PF	26-FEB-1999; 99WO-US04316.
XX	
PR	20-MAR-1998; 98US-0045284.
PR	12-NOV-1998; 98US-0190911.
XX	
PA	(REGC) UNIV CALIFORNIA.
PA	(SYNT) SYNTEX USA INC.
XX	
PI	Blstrup A, Rosen SD, Tangemann K, Hemmerich S;
XX	
DR	WPI; 1999-580442/49.
DR	P-PSDB; AAY39919.
XX	
PT	Human and murine glycosyl sulfotransferase 3 and related
PT	polynucleotides
XX	
PS	Claim 4; Fig 3; 59pp; English.
XX	
CC	This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of

the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues. The sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.

SQ Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;

Query Match 30.2%; Score 354; DB 20; Length 1926;
Best Local Similarity 60.9%; Pred. No. 1.9e-59;
Matches 614; Conservative 0; Mismatches 385; Indels 9; Gaps 2.

QY	124	gtgcagctgctgctgctgctgctggcctcggcctcgaacatctcttcttgccagactcttc	183
Db	528	gtcagtgtgctgctgctctctgctgctgagcaggatacatcttttgtggcaagatttc	587
QY	184	agccagcacccccaagctctctaccctgatggagccgcgtgccatgtgtggaccaacctg	243
Db	588	ggcgagcaccccgatgtgttctacctgatggagcctgcctggcatgtgtggatgacttcc	647
QY	244	tgcagggcagcgcggcgaacgctgcacatggccgtgcgcacctgatgcgctctacttt	303
Db	648	accagcagcacagccctggaaagctgcacaTggctgtggggatctctcgtggttcccgtcttc	707
QY	304	ttgtgcacatgaagctgtgtttgatgcctacatg--ccacagagccgaacactgttcoooc	360
Db	708	ctgtgtacatgagctgtttgatgctacatgaaccagggcccccgaaaacagtcacag	767
QY	361	ttttcaactgggcaacagcgcgcgtgtgtctgcgcgcgcgcctgcagcctttccc	420
Db	768	ctcttcagtgggagcaaacgcggccctgtgctcagcgcctgtgtgtgacttctccct	827
QY	421	cggaggcacatatcagaagcaggagctatgcgaagacaactgtgcacgcgcgcacattcagc	480
Db	828	gccacagagatcagctcacccaaagcactgcaaagctgctcggcggtcagcagcccttgat	887
QY	481	ctgcccggaggcctgcgcctcctacagccagctgtgtctcaaggaggtgcgctcttc	540
Db	888	atggtggaaagccctgcgcctcagcgtctgtgttactcaaggaggtgcgtttcttc	947
QY	541	aacctgcagtgctctacccgcgtgtcagcagccccgcgcctaaccctgcgactcgtcac	600
Db	948	agcctgcaggccoctctatccactactcaccgaccccttccctcaacctgcacgtgcagc	1007
QY	601	ctggtgcgcgacccgggcgcgtgtcgtcgtctccgggaggcggcgccgacatactgcga	660
Db	1008	ctggtccgagaccccccggccgtgttctcgatccccgggagcacaccacatagaactcgt	1067
QY	661	cgcgcacaacgycatcgtgtgtggccaccaacgcgaagctgggtggaggccgacctcaactg	720
Db	1068	gttgacagtcataatgtgtcagggcagcatttggaaaaagcataaggagggaagaccagccc	1127
QY	721	cgcctgattgcgaggtgtgcgcgacccaactgcgcctcgcgagcgccgcacacactcaag	780
Db	1128	tattatgcattgaagatcatctgcaaaagccagggtggacatagtcaaggccatccaaacc	1187
QY	781	cgcgcaccccttctgcgggcgcgtacacgcctggtgtgccttcgaggacactggcgaggag	840
Db	1188	ctccctgaagctctgcagcagcgtacctgttccctgagtgatgaggacctggttcgggca	1247
QY	841	ccgctggcagagatccgcgcactcttacgccttcacgccttcacgcctcagccacacagctc	900
Db	1248	ccctggcccgacgacaccagacataataaatgtgtggggttggattttttggccacactc	1307
QY	901	gagggcctggatccacaacatcacccacggctcggggatctgcgaagccaatcgagcccttc	960
Db	1308	caaacatgggtttacaatgtcacccgcggcaaggagcattgggtcagc-----atgccttc	1365
QY	961	cataactccttaggaatgcgcgacagctctctccaggcctggcgccacgcgttgcccttc	1020

[illegible]

RESULT 14

AAK91803
ID AAK91803 standard; cDNA: 877 bp.

XX	
AC	AAK91803;
XX	
DT	06-NOV-2001 (first entry)

XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 263.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

OS Homo sapiens.

XX
PN
EP1130094-A2.

XX 05-SEP-2007

XX
07-THF-3000.XX
08 - TTTT - 1000 -

PR 11-JAN-2000;

XX

XX
XX

PI Wakamatsu A,

WPI; 2001-524

XX
PT
830 Primers u

XX use in genetic

PS
XX
Claim 2; SEQ

CC The invention

isolated and

CC cDNA are used

length enrichment

CC without any s.

Note: The sequence of

CC
XX
specification

Sequence 877 : SQ

Quarry Master

Best Local Similarity

100

0 —
 7 —
 8 —
 9 —
 6 —
 5 —
 4 —
 3 —
 2 —
 1 —
 0 —
 7 —
 8 —
 9 —
 6 —
 5 —
 4 —
 3 —
 2 —
 1 —
 0 —

on day 60

QY I/8 ctcttcag

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 19:02:45 ; Search time 1398.08 Seconds
(without alignments)
2510.138 Million cell updates/sec

Title: US-09-593-828-3
Perfect score: 2044
Sequence: 1 ggctcaggtccactgtgtc.....aaaaaaaaaaaaaaaaaaaaa 2044

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_032802.*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2044	100.0	2044	22	Human glycosyl sul
2	1973	96.5	2170	22	Human glycosyl sul
3	1827.2	89.4	2988	21	Human ORFX ORF1711
4	1753.6	85.8	160552	22	Human glycosyl sul
5	1111.4	54.4	1694	22	Human glycosyl sul
6	947	46.3	1647	24	Human drug metabol
7	751.2	36.8	1989	22	Mouse glycosyl sul
8	750	36.7	1937	24	Murine intestinal-
9	400	19.6	1979	22	Human full-length

10	400	19.6	2032	20	AAZ20792	Human glycosyl sul
11	398.4	19.5	1333	24	AA316947	Human L-selectin s
12	377	18.4	2065	21	AA294211	Human transferrase
13	354	17.3	1926	20	AAZ20793	Mouse glycosyl sul
14	297.6	14.6	877	22	AAK91803	Human cDNA 5'-end
15	297.6	14.6	877	22	AAK91803	Human cDNA clone r
16	235.2	11.5	2156	19	AAK93921	Glycosaminoglycan
17	222	10.9	12503	22	AAV21200	Human DNA for a no
18	216.4	10.6	107820	22	AA331470	Human ATP-binding
19	214.8	10.5	32249	22	AA316230	Human nervous syst
20	211.8	10.4	24843	24	AA317155	Human genomic DNA
21	211	10.3	1043	21	AA317764	Human interleukin-
22	211	10.3	1601	21	AA314459	Human low adenosin
23	211	10.3	1601	21	AA315191	Human adenosine re
24	211	10.3	1601	21	AA315191	Human adenosine re
25	211	10.3	1618	12	AA46814	Interleukin-10 (IL
26	211	10.3	1618	12	AA46814	PH15C insert conta
27	211	10.3	1618	14	AA46814	Human cytokine syn
28	211	10.3	1618	21	AA46814	Human CSIF coding
29	211	10.3	1618	21	AA46814	Human CSIF protein
30	211	10.3	1618	22	AA46814	Human CSIF CDNA.
31	211	10.3	1645	21	AA46814	Human CSIF CDNA.
32	211	10.3	2797	22	AAH18187	Human CSIF CDNA.
33	211	10.3	15630	21	AA213117	Human CSIF CDNA.
34	211	10.3	15630	21	AA213117	Human CSIF CDNA.
35	209.8	10.3	32152	22	AA35195	Human CSIF CDNA.
36	209.2	10.2	23989	22	AA35195	Human CSIF CDNA.
37	209	10.2	14209	22	AAK89137	Human CSIF CDNA.
38	208.8	10.2	21477	22	AAK89137	Human CSIF CDNA.
39	208.8	10.2	21480	22	AAK89137	Human CSIF CDNA.
40	208.8	10.2	27571	22	AAK89137	Human CSIF CDNA.
41	208.8	10.2	27571	22	AAK89137	Human CSIF CDNA.
42	208.4	10.2	15589	22	AAK89137	Human CSIF CDNA.
43	208.4	10.2	15610	22	AAK89137	Human CSIF CDNA.
44	208.4	10.2	15610	22	AAK89137	Human CSIF CDNA.
45	208.2	10.2	4125	22	AAK84260	Human CSIF CDNA.

ALIGNMENTS

RESULT	1
AAAD02699	
ID	AAAD02699 standard; cDNA; 2044 BP.
XX	AAAD02699;
AC	AAAD02699;
XX	02-MAY-2001 (first entry)
DT	Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.
XX	Human glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
DE	therapy; selectin binding inhibitor; gene therapy; inflammation;
XX	systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW	polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW	glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenalitis;
KW	Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
KW	demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW	myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
KW	asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW	chromosome 16q23.1; ss.
XX	Homo sapiens.
OS	
XX	Key
XX	Location/Qualifiers
FT	218..1390
FT	/tag= a
FT	/product= "Human glycosyl sulfotransferase-4alpha
FT	(GST-4alpha)"
FT	/note= "CDS is specifically claimed as SEQ ID NO: 4
FT	in claim 6 (page no: 41) of the specification"
XX	
XX	W0200106015-A1.

QY	1501	gggagtttggggtccctcccctgaagtaagcaaggaactgaacgtttctttctctccgtatt	1560
Db	1501		
QY	1561	ctcggtttcccttggagtctcttgagctgcctctctcatcagtgcaactcttcattgaaa	1620
Db	1561		
QY	1621	agcaactcttgccccctaccctctcttcctggcgcgaggagtaagttaactgctaaataaatta	1680
Db	1621		
QY	1681	aatgtgtgcagcgcggtgcgtggtcctcatgcctgtaatccagcattttgaagagctcg	1740
Db	1681		
QY	1741	aggcgggtgatacaactgaaggtcaaggattcaaaacacgctctgccaacatagtgaaacccc	1800
Db	1741		
QY	1801	cctctctactaaaatgcataaatttagtcccgcgtggtggcacaactcctgtaatccacgc	1860
Db	1801		
QY	1861	tacttagaggtcgagtgaggaaatacacttgactccaaaggtggaggttcagctaagc	1920
Db	1861		
QY	1921	tgaatacatgccactgcacctagcttggttgccaagcaaacctctctcaaaaaataa	1980
Db	1921		
QY	1981	ttataaatttgttcaaaagtcctgcgcgaataaaaaaataaaaaaataaaaaa	2040
Db	1981		
QY	2041	aaaa 2044	
Db	2041	aaaa 2044	

RESULT 2

AAAD02698
ID AAAD02698 standard; cDNA; 2170 BP.

AA
AC AAD02698;

DT 02-MAY-2001 (first entry)

Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.

Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
therapy; selectin binding inhibitor; gene therapy; inflammation;
systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
glomerulonephritis; myasthenia gravis; Sjogren's syndrome; dermatitis;
Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
asthma; hypersensitivity; rheumatic fever; tissue rejection;
chromosome 16q23.1; ss.

OS Homo sapiens.

Key	Location/Qualifiers
FH	9.343
FT	

```
FT      /*tag= a
```

```

ET
...
/*tag= b

```

```
FT /product= "Human glycosyl sulfotransferase-4alpha  
FT (GST-4alpha)"
```

FT /note= "CDS is specifically claimed as SEC ID NO. 4

FT		
FT	3'UTR	in claim 6 (page no: 41) of the specification"
FT		1517..2134
FT		/*tag= C
PN	WO200106015-A1.	
XX	25-JAN-2001.	
XX	19-JUL-2000; 2000WO-US19741.	
XX	20-JUL-1999; 99US-0144694.	
PR	13-JUL-2000; 2000US-0593828.	
XX		
PA	(REGC) UNIV CALIFORNIA.	
XX		
PI	Rosen SD, Lee JK, Hemmerich S;	
DR	WPI: 2001-138471/14.	
DR	P-PSDB; AAY72639.	
XX		
PT	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for	
PT	diagnostic and therapeutic agent screening applications -	
PS	Claim 6; Page 62; 128pp; English.	
XX		
CC	The present sequence is human glycosyl sulfotransferase-4alpha (GST-4	
CC	alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.	
CC	GST is a type 2 membrane protein useful for inhibiting a binding event	
CC	between a selectin and a selectin ligand, which comprises contacting the	
CC	selectin with a non-sulphated selectin ligand, GST and a small molecular	
CC	agent that inhibits the sulphation activity of GST. GST is also useful	
CC	in inhibiting a selectin mediated binding event. GST is useful in gene	
CC	therapy to treat disorders such as acute or chronic inflammation,	
CC	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis	
CC	nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,	
CC	glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's	
CC	disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious	
CC	anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,	
CC	dermatitis, myocarditis, regional enteritis, adult respiratory distress	
CC	syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,	
CC	bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection	
CC	during transplantation.	
XX		
SQ	Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;	

```

Query Match          96.5%;   Score 1973;   DB 22;   Length 2170;
Best Local Similarity 99.7%;   Pred. No. 0;
Matches 1998; Conservative 0; Mismatches 5; Indels 2; Gaps 2

```

QY	41	gaaagctcccaagt	gagaaactggctctctggagactctgtgtgcatagagtattcaac	100
Db	167	gaaagctcccaagt	gagaaactggctctctggagactctgtgtgcatagagtattcaac	226
QY	101	caccttaagaagacct	gtgcttccctggaaacacagatgtcgagacatctcccatgatt	160
Db	227	caccttaagaagacct	gtgcttccctggaaacacagatgtcgagacatctcccatgatt	286
QY	161	tgtgatcagcgtgtg	cagctctccacagacgcttgacgctggccccacgcccgcctg	220
Db	287	tgtgatcagcgtgtg	cagctctccacagacgcttgacgctggccccacgcccgcctg	346
QY	221	tggctgccacggtt	ctccagcaagacagtgacagtctctctctggacacacacctgc	280
Db	347	tggctgccacggtt	ctccagcaagacagtgacagtctctctctggacacacacctgc	406
QY	281	ctctcgtcttcaat	atctcccgccagggccctcatccccacggcgccgagagatcgt	340
Db	407	ctctcgtcttcaat	atctcccgccagggccctcatccccacggcgccgagagatcgt	466
QY	341	gtgcaacgtgctg	ctctcgtggcgctcgggctcatctctttggccagctcttc	400
Db	467	gtgcaacgtgctg	ctctcgtggcgctcgggctcatctctttggccagctcttc	526

QY	1481	a-caggctccactataaaccggaggtttggggtcctccctgaagtaagcaggactgc	1539
Db	1607	agcaggctccactataaaccggaggtttggggtcctccctgaagtaagcaggactgc	1666
QY	1540	acgtttctctctcctcattcctcgttttctcttctgagctctctgagctccttctcat	1599
Db	1667	acgtttctctctcctcattcctcgttttctcttctgagctccttctgagctccttctcat	1726
QY	1600	caggctgactctctcctgaaagcaactcttgcctcactccttctcctgagcaggagta	1659
Db	1727	caggctgactctctcctgaaagcaactcttgcctcactccttctcctgagcaggagta	1785
QY	1660	agttactgtaataataataatgtgtccaggccgggtgcgtgtgctcagctcgttaa	1719
Db	1786	agttactgtaataataataatgtgtccaggccgggtgcgtgtgctcagctcgttaa	1845
QY	1720	tcaccagcttttgagagctgagcggtgtggtacactgaggtcaggtattcaaaaccgc	1779
Db	1845	tcaccagcttttgagagctgagcggtgtggtacactgaggtcaggtattcaaaaccgc	1905
QY	1780	ctggcccaacatagtgaaacccctctctactaaaaatgaaaaatagtcgcgctggtg	1839
Db	1906	ctggcccaacatagtgaaacccctctctactaaaaatgaaaaatagtcgcgctggtg	1965
QY	1840	gcacactcctgtaataccagctacttagagctgaggtgtgggaaatcacttggactcca	1899
Db	1966	gcacactcctgtaataccagctacttagagctgaggtgtgggaaatcacttggactcca	2025
QY	1900	aaggtggaggttgagtaagctgaaatcatgccactgcacccctagcttggtggcaagc	1959
Db	2026	aaggtggaggttgagtaagctgaaatcatgccactgcacccctagcttggtggcaagc	2085
QY	1960	aaaactctatcaaaaaataataataattgttcaaaagctcctgcgaaaaaaaaa	2019
Db	2086	aaaactctatcaaaaaataataataattgttcaaaagctcctgcgaaaaaaaaa	2145
QY	2020	aaaaaaaaaaaaaaaaaaaaa	2044
Db	2146	aaaaaaaaaaaaaaaaaaaaa	2170
RESULT	3		
AC	AAC76156		
DT	08-FEB-2001	(first entry)	
DE	Human ORF1711	polynucleotide sequence SEQ ID NO:3421.	
XX	Human; open reading frame; ORF; detection; cytotropic; hepatotropic;		
KW	vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;		
KW	anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;		
KW	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KW	hypotensive; dermatological; immunosuppressive; antiinflammatory;		
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
KW	thrombosis; contraceptive; ss.		
OS	Homo sapiens.		
XX	WO200058473-A2.		
PN	05-OCT-2000.		
PD	31-MAR-2000;	2000WO-US08621.	
XX			
PF			

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications -
PS Example 1; Page 62-104; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4 (GST-4) genomic
CC DNA encoding GST-4alpha and GST-4beta. GST-4 gene is found on
CC chromosome 16q23.1.
CC GST is a type 2 membrane protein useful for inhibiting a binding event
CC between a selectin and a selectin ligand, which comprises contacting the
CC selectin with a non-sulphated selectin ligand, GST and a small molecular
CC agent that inhibits the sulphonation activity of GST. GST is also useful
CC in inhibiting a selectin mediated binding event. GST is useful in gene
CC therapy to treat disorders such as acute or chronic inflammation,
CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.

XX Sequence 160552 BP; 40281 A; 37573 C; 38015 G; 44564 T; 119 other;

Query Match 85.88; Score 1753.6; DB 22; Length 160552;
Best Local Similarity 99.18; Pred. No. 4.2e-280;
Matches 1795; Conservative 0; Mismatches 14; Indels 3; Gaps 3;
QY 197 cgtgtgccccagccgcccgcattggtgccacggttctccagcaagacagtgacagtg 256
DB 47935 cgaaggccccagccgcccgcattggtgccaggttctccagcaagacagtgacagtg 47994
QY 257 ctctctctggcacagacacactgctctctctctctctctctctctctctctctctc 316
DB 47995 ctctctctggcacagacacactgctctctctctctctctctctctctctctctc 48054
QY 317 tcccacagccggcgagagatggtgcacgtgctgctgctgctgctgctgctgctgctg 376
DB 48055 tcccacagccggcgagagatggtgcacgtgctgctgctgctgctgctgctgctg 48114
QY 377 toatctcttgggccaagtctcttaagccagcaaccccgacgtctcttaccctgatgagccc 436
DB 48115 toatctcttgggccaagtctcttaagccagcaaccccgacgtctcttaccctgatgagccc 48174
QY 437 gcgtggcatgtggaccacccgttcacagggcagcgcgcgaacgcgtgcacatggccgtg 496
DB 48175 gcgtggcatgtggaccacccgttcacagggcagcgcgcgaacgcgtgcacatggccgtg 48234
QY 497 cgcgacctgatgcgtctatctttttgtgcgacatgacgtgtttgatgctacatgcc 556
DB 48235 cgcgacctgatgcgtctatctttttgtgcgacatgacgtgtttgatgctacatgcc 48294
QY 557 cagagcgaacactgtccgccttttccaaactggcgaacgagcgcgcgtgtgctgcgcg 616
DB 48295 cagagcgaacactgtccgccttttccaaactggcgaacgagcgcgcgtgtgctgcgcg 48354
QY 617 ccgcctcagacgcctttcccccaggaacacatcagcaagcagcagctatgcaagacactg 676
DB 48355 ccgcctcagacgcctttcccccaggaacacatcagcaagcagcagctatgcaagacactg 48414
QY 677 tgcacgcgcgagccattcagccctggcccggaggccctgcgcctcctacacccacgtgtg 736
DB 48415 tgcacgcgcgagccattcagccctggcccggaggccctgcgcctcctacacccacgtgtg 48474
QY 737 ctcaaggagtgcgcttcttcaacctgcagtgctctctacccgctgctcagcgaccgcgcg 796
DB 48475 ctcaaggagtgcgcttcttcaacctgcagtgctctctacccgctgctcagcgaccgcgcg 48534
QY 797 ctcaacctgcgcatctgcacctggtgcgcgaccccgggcgctgctgcgctcccggag 856
DB 49614 agtggggaatacacttggactccagaggtggagggttgagtaagctgagatcatgccac 49673

DB 48535 ctcaacctgcgcatcgtgcacctggtgcgcgaccccgggccgctgctgcgtccccggag 48594
QY 857 gcggcgggcccgatactggtcacgcgcacacggcgcattcgtctggtggcaccacggcgaagtgg 916
DB 48595 gcggcgggcccgatactggtcacgcgcacacggcgcattcgtctggtggcaccacggcgaagtgg 48654
QY 917 gtggaggcgacccctcaactgcctgcctgatttcgcgaggtgtgctgcgacgacactgcgcate 976
DB 48655 gtggaggcgacccctcaactgcctgcctgatttcgcgaggtgtgctgcgacgacactgcgcate 48714
QY 977 gcggaggcgccacactcaagccgcacacttccctgcgcggcggtacacgcctggtgcgc 1036
DB 48715 gcggaggcgccacactcaagccgcacacttccctgcgcggcggtacacgcctggtgcgc 48774
QY 1037 ttcgaggacctggcggggagccgctggcagagatcccggcacttaagccttccaccgcg 1096
DB 48775 ttcgaggacctggcggggagccgctggcagagatcccggcacttaagccttccaccgcg 48834
QY 1097 ctgacctcagccacacagctcgagccctggatccacacacatcaccacagggctcggggac 1156
DB 48835 ctgacctcagccacacagctcgagccctggatccacacacatcaccacagggctcggggac 48894
QY 1157 ggcaagccaactcgagccctccactctcctaggaatcgcgcaacgctctccacagcc 1216
DB 48895 ggcaagccaactcgagccctccactctcctaggaatcgcgcaacgctctccacagcc 48954
QY 1217 tggcgccacgcgttgcccttccactaagatcctgcgtgcgaggaggtgctgcgcgcgcgcg 1276
DB 48955 tggcgccacgcgttgcccttccactaagatcctgcgtgcgaggaggtgctgcgcgcgcgcg 49014
QY 1277 ctgcagctgcctggactacccgctgtgtactctgcggcaccagcagctgacactaccctg 1336
DB 49015 ctgcagctgcctggactacccgctgtgtactctgcggcaccagcagctgacactaccctg 49074
QY 1337 gatctggtgcctgcacagggcccaacactcagctgggcacgcctgactgcactgagaactc 1396
DB 49075 gatctggtgcctgcacagggcccaacactcagctgggcacgcctgactgcactgagaactc 49134
QY 1397 tggccttagagcgaagccccgaactgtggtgcgcagggcccagggaagcgactgcattgggtg 1456
DB 49135 tggccttagagcgaagccccgaactgtggtgcgcagggcccagggaagcgactgcattgggtg 49194
QY 1457 aaaaggagctggggcgcatggggaa-caggtccctactatcaaccggaggtttggggtcc 1515
DB 49195 agagggagctggggcgcatggggagcaggtccctactatcaaccggaggtttggggtcc 49254
QY 1516 tccctgaagtgaagcgaaggactgcacgtttcttctcctcctgattctcgtgttccctttg 1575
DB 49255 tccctgaagtgaagcgaaggactgcacgtttcttctcctcctgattctcgtgttccctttg 49314
QY 1576 agtcttctggagctgccttctcactcaggtgcacttctcatttgaaagaaacactcttgcgcc 1635
DB 49315 agtcttctggagctgccttctcactcaggtgcacttctcatttgaaagaaacactcttgcgcc 49373
QY 1636 tacctcttctggcgagcaggtgaagttaactgctataataataataatgctgacagcc 1695
DB 49374 tacctcttctggcgagcaggtgaagttaactgctataataataataatgctgacagcc 49433
QY 1696 ggggtgcgttgctcactgcctgtaatcccgacattttgagaggtgcagcggttgatcac 1755
DB 49434 ggggtgcgttgctcactgcctgtaatcccgacattttgagaggtgcagcggttgatcac 49493
QY 1756 ctgaggtcagga-ttcaaaaccagcctggccaacatagtgaaacccctctctactaaaa 1814
DB 49494 ctgaggtcagaggttgaaaccagcctggccaacatagtgaaacccctctctactaaaa 49553
QY 1815 atcaaaaaattagtcggcggtggtggcacactcctgtaatccacagctacttagagctg 1874
DB 49554 atcaaaaaattagtcggcggtggtggcacactcctgtaatccacagctacttagagctg 49613
QY 1875 agtgggaaatacacttggactccaaaggtggaggttgagtaagctgaaatcatgccac 1934
DB 49614 agtgggaaatacacttggactccaaaggtggaggttgagtaagctgaaatcatgccac 49673

[illegible]

RESULT

RESOLUT 0
AAD24670
ID AAD24670 standard; cDNA; 1647 BP.
.....

AC AAD24670:

DT 12-MAR-2002 (first entry)

Human drug metabolising enzyme (DME)-5 cDNA.

XX	Human; drug metabolising enzyme; gene therapy; autoimmune disorder;
KW	inflammatory disorder; acquired immune deficiency syndrome; infection;
KW	AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW	allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW	cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
KW	gastrointestinal disorder; metabolic disorder; developmental disorder;
KW	liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW	goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW	DME-5; ss.

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

```
CDS
1..1188
/*tag= a
/product= "Human DME-5 protein"
1..96
/*tag= b
97..1185
/*tag= c
/product= "Human mature DME-5 protein #1"
1..105
/*tag= d
106..1185
/*tag= e
/product= "Human mature DME-5 protein #2"
```

WO200179468-A2.

25-OCT-2001.

12-APR-2001; 2001WO-US11869.

13-APR-2000; 2000US-197590P.

15-APR-2000; 2000US-198403P;
28-APR-2000; 2000US-200185P;

05-MAY-2000; 2000US-202234P.

11-MAY-2000; 2000US-203509P.

(INCY -) INCYTE GENOMICS INC.

Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM; Yao MG, Yue H, Tang YR, Patterson C, Das D, Sanjanwala MS; Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA; Au-Yong J;

WPI; 2002-066363/09.

P-PSDB; AAEl5438.

Novel isolated human drug metabolizing enzymes referred as DME 1-10, useful for diagnosing, treating, or preventing disorders associated with aberrant expression of DME such as allergy, anemia, asthma, infertility -

Claim 5; Page 139; 143pp; English.

The invention relates to human drug metabolising enzymes referred as DME and nucleic acid molecules encoding such enzymes. Polynucleotides of the invention are useful for assessing toxicity of test compounds and in gene therapy. Sequences of the invention are useful in the diagnosis, prevention and treatment of autoimmune/inflammatory disorders, such as acquired immune deficiency syndrome (AIDS), adult respiratory distress syndrome, allergies, anaemia, atherosclerosis, asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis, uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic infections; cell proliferative disorders such as actinic keratosis, atherosclerosis, atherosclerosis, Duchenne and Becker dystrophy, hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome, epilepsy; endocrine disorders such as disorders of the hypothalamus and pituitary resulting from lesions such as primary brain tumours, adenomas, infarction associated with pregnancy, aneurysms, vascular malformations; eye disorders such as conjunctivitis, iritis, retinitis, glaucoma, pigmentosa; metabolic disorders such as Addison's disease, cystic fibrosis, diabetes, goitre, glycogen storage diseases, hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies, Wernkes syndrome, mannosidosis, obesity; gastrointestinal disorders such as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis, hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental disorders. The present sequence is human DME-5 protein cDNA.

Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

CC The present sequence is mouse glycosyl sulfotransferase-4 (GST-4) cDNA.
 CC GST-4 gene is found on chromosome 8E1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenalitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 XX

Sequence 1989 BP; 415 A; 586 C; 550 G; 438 T; 0 other;

Query Match 36.8%; Score 751.2; DB 22; Length 1989;
 Best Local Similarity 70.9%; Pred. No. 3.6e-115;
 Matches 1046; Conservative 0; Mismatches 413; Indels 17; Gaps 3;

QY 202 gcccccagccgcccgcagatgtgtgtccacaggtttctccagcagacagtgacagtctctc 261
 DB 91 ggcctagccagagtgatgcgtaccocctttctccagcagactgtcatgtctctc 150
 QY 262 cctggcacaacaccctgcctctctctctctctctctctctctctctctctctctctc 321
 DB 151 gatgttagcag--actggcatctctgtctctctctctctctctctctctctctctc 207
 QY 322 agccggcggcagagatgtgtgcacgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 381
 DB 208 agcaggccttgggagcgtgtgcagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgt 267
 QY 382 atttttggccagctctctccagcagcccccagcagctctctctctctctctctctct 441
 DB 268 ctctctggcagctctctccagcagcccccagcagctctctctctctctctctctct 327
 QY 442 gcatgtgtgacacacccctgtgcagcagcagcagcagcagcagcagcagcagcagc 501
 DB 328 gcaogtctggatcgtgtgcagcagcagcagcagcagcagcagcagcagcagcagc 387
 QY 502 cctgatgcgct 561
 DB 388 cctgatgcgct 447
 QY 562 ccgaacacctctcgcct 621
 DB 448 ccgaacacctctcgcct 507
 QY 622 ctgcagcgcctctctcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 681
 DB 508 ctgcagcgcctctcgcagcagcagcagcagcagcagcagcagcagcagcagcagc 567
 QY 682 ggcggcagcattcagcctcgcggcggcggcggcggcggcggcggcggcggcggcggc 741
 DB 568 aacgcgccctcgcctcgcagcagcagcagcagcagcagcagcagcagcagcagc 627
 QY 742 gaaggtgcctctctcagcagcagcagcagcagcagcagcagcagcagcagcagc 801
 DB 628 ggaaggtgcctctctcagcagcagcagcagcagcagcagcagcagcagcagcagc 687
 QY 802 cctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 861
 DB 688 cctgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 747
 QY 862 gggcccgatctgacgcagcagcagcagcagcagcagcagcagcagcagcagcagc 921
 DB 748 agccaaggcgcgtgcagcagcagcagcagcagcagcagcagcagcagcagcagc 807

QY 922 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 981
 DB 808 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 867
 QY 982 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 1041
 DB 868 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 927
 QY 1042 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 1101
 DB 928 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 987
 QY 1102 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1161
 DB 988 tctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1047
 QY 1162 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 1221
 DB 1048 ggcgcacctcaactgcgcgtgattcgcaggtgtgcgcagcagcagcagcagcagc 1107
 QY 1222 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1281
 DB 1108 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1167
 QY 1282 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1341
 DB 1168 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1227
 QY 1342 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1401
 DB 1228 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1287
 QY 1402 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1461
 DB 1288 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1344
 QY 1462 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1521
 DB 1345 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1404
 QY 1522 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1570
 DB 1405 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1464
 QY 1571 cctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1630
 DB 1465 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1524
 QY 1631 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1666
 DB 1525 gctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1560

RESULT 8

AAS16948

ID AAS16948 standard; cDNA; 1937 BP.

XX AAS16948;

AC AAS16948;

XX AAS16948;

DT 12-MAR-2002 (first entry)

XX Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) cDNA.

DE Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) cDNA.

XX Mouse; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;

KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;

KW ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;

KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;

KW delayed-type hypersensitivity reaction; hyperplastic thymus; antiulcer;

KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;

KW antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.

XX Mus musculus.

OS Mus musculus.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EP1130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-0114089.

08-JUL-1999; 99JP-0194486.

11-JAN-2000; 2000JP-0118774.

02-MAY-2000; 2000JP-0183765.

(HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y; Wakamatsu A, Sudiyma T, Nagai K, Kojima S, Otsuki T, Koga H; WPI; 2001-524255/58.

P-PSDB; AAM93309.

830 Primers useful for synthesizing full length cDNA clones and their use in genetic manipulation -

Claim 8; SEQ ID NO 2816; 1380pp + sequence listing; English.

The invention relates to primers for synthesizing full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' and 3'-ends of the cDNA molecules have been determined. Primers for synthesizing the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a full length human cDNA of the invention.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.

Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 19.6%; Score 400; DB 22; Length 1979;

Best Local Similarity 63.6%; Pred. No. 2.4e-57;

Matches 645; Conservative 0; Mismatches 360; Indels 9; Gaps 2

QY 335 gatcgtgtgcacgtgctgtgctgtctctcgtggcgctggcgctacatctcttcttggcgacg 394

Db 189 gagcgatgcacgtgctgtgttctgtctctcgtggcgctgtgctctcttcttggcgacg 248

QY 395 ctcttcagccagcaccgccagctctctacctgatggcgccgctggcgctgtggtgacg 454

Db 249 ctttttggcgagcaccagatgtttctacctgatggcgccgctggcgctgtggtgacg 308

QY 455 accctgtcgcaggcagcgcggcgcaacgtgcacatgcccgtgcgcacctgatgcgtct 514

Db 309 acctcgaagcagagcgcgcgtggtgatgctgcacatgctgtgcgggatctgatacggcc 368

QY 515 atcttttggcacatggcgtgtttgatgctcctacatgcacag--ccgaaacctg 571

Db 369 gtcttgtgcacatgagcgtcttga-gcctacatggaacctgtgcccgagacag 428

QY 572 tccgcttttcaactggcgaaagcgcgcgcgtgtgtcgcgcgcgcctgcagccc 631

Db 429 tccagcctcttcagtgggagaacagcgggacctgtgtctgcacctgcctgtgacatc 488

QY 632 ttccccgagaccatcagcaagcaggacgctatgaagacactgtgcagcggcagcca 691

Db 489 atcccacaagatgaatcattccccgggctcactgcagcgtcctgtcagtcacacgcc 548

QY 692 ttcagcctggccggaggcctaccgtctctacagcacgtgtgtctcaaggaggtgcgc 751

betal,3 α NT that directs expression of a MECA-79 antigen. This can be done by administering to the subject an oligosaccharide L-selectin antagonist that inhibits binding of L-selectin to a MECA-79 antigen, for example by administering antibody material that specifically binds betal,3 α NT, and/or a betal,3 α NT antisense nucleic acid molecule. L-selectin sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression. Alternatively, the expression or activity of LST-2 or its active fragment can be reduced in combination with reducing the expression or activity of betal,3 α NT. The method is useful for treating L-selectin mediated conditions such as Crohn's disease and ulcerative colitis, inflammatory disorders of the skin such as allergic contact dermatitis, psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type hypersensitivity reactions, diabetes and hyperplastic thymus. This sequence represents cDNA encoding human LST-2.

Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 19.5%; Score 398.4; DB 24; Length 1333;
Best Local Similarity 63.5%; Pred. NO. 4.3e-57;
Matches 644; Conservative 0. Mismatches 261.

2:

QY	335	gagctgtgtgcacgtgctgtgtcgttcctcgttgccgtcgggctcaaccttcttttggccag	394
Dd	213	gagcgatgaactgtgtgttcgtctcctggcgtctgtctcttttgttggggcag	272
QY	395	ctcttcaggcacaccacgactcttctacctgatggagccgcgtggcatgttgacc	454
Dd	273	cttttggcagcacccagatgtttctacctgatggagccgcgtggcagtgtggatg	332
QY	455	acctgtcgagggcagcgcggaacgcgtgcacatggccgtgcggacotgatgcctct	514
Dd	333	acctcaagcagcacccgcctggatcgtgcacatggcgtgcgggatctgatcagggcc	392
QY	515	atctttttgtcgacaatgaactgtttgatcctacatgccacag---ccgaacctg	571
Dd	393	gtctcttgtgcacatgagcgtctttgatcctacatggaaacctggctcccgagacag	452
QY	572	tccgccttttcaactgggccaagcgcgcgtgtgtctgcgcgcgcgcctgcagccc	631
Dd	453	tccagcctcttcagtgggagacagccggccctgtgttctgcacctgctgtgacac	512
QY	632	tttccccaggcacocatcagcaagaagcagcttatcaagacactgtgcacgcgcgaccca	691
Dd	513	atcccacaagatgaatacatcccgcggctcactgcaggtcctgtgcagtaaacagccc	572
QY	692	ttcagccttggccggaggcctgcgcctctacagccacgttggtgtcacaaggatggc	751
Dd	573	tttggagggtggagaagcctgcgcctcctacagccacgtggtgtcaaaggatggc	632
QY	752	ttcttcaacctgcaggtgtcttacccgtgtctcagcagccgcctcaactgcagctc	811
Dd	633	ttcttcaacctgcagtccctctaccgcgtctgaagaccctcctcaactcaactgacac	692
QY	812	gtgcacctgtgtcgacaccgcggccgtgtcgtcgtcccgaggagcggcgccgata	871
Dd	693	gtgcacctgttccgggaccccccggcgtgttcggttcccgagAACGCaacaaggagat	752
QY	872	ctggcagcgcacaacgcctcgtctaggcaaccaacggcaagtgggtggagccgacccct	931
Dd	753	ctcatgattgaactgcgcatgtgtatggggcagcatgagcaaaactcaagaaggagac	812
QY	932	cacctgcgcctgattcgcagaggtgtgcgcagcaactggcctatgcgcgagccgcacca	991
Dd	813	caacctactatgtgatcaggtcatctgcacaagccagctggagatctcaagaccatc	872
QY	992	ctcaagcgcacaacctctctgcgcgcgtacccgctgtgtgcgtctcgaggacctggcg	1051
Dd	873	cagtccttgcacaagccctgcaggacgctaactgctgttgcgtatgaggacctggct	932
QY	1052	cgggagccgctggcagagatccgcgcactatcagccttcaaccgctgacccctcaogcca	1111
Dd	933	cgaacccctgtggcccgaacttccgaatatgaaatctcgtgggattggaaacttqccc	992

Qy	1112	cagctcgaggcctggatccacaacatccaccacgggtcgggatcggaagccaatcgag	1171
Db	993	catcttcagaccctgggtgataacatcccgaggaaggcattgggtga-----ccac	1046
Qy	1172	gccttcatacttcgctcaggaaatcgcgcaacggtctccagagcctgcgcgaacgcgttg	1231
Db	1047	gctttccacacaaatgccagggaatgcccttaatgctccagaggtggcgtgctcttg	1106
Qy	1322	cccttctaactaagatcctgcgcgtcgaggaaggtgtgcgcggcgcgctcagcctcgtggc	1291
Db	1107	ccctatgaagaagtttctcgacttcagaagaacgctgtgcgatgcgaatattcctgggc	1166
Qy	1292	tacggccgtgtgaactctcggaaccagacgtagcctcaacctggatcctgggtg	1345
Db	1167	taacggcacgctcgactgaacaaagaagaacaaacctgttgctggatctctg	1220

RESULT 12

AAZ94211
ID AAZ94211 standard; cDNA; 2065 BP.

AC AAZ94211;

DT 19-JUN-2000 (first entry)

Human transferase TRNSFS-11 cDNA clone 2617407CBL.

Transferase; TRNPSF5-11; human; antitumour; cell proliferation;
inflammation; gastrointestinal disorder; developmental disorder;
genetic disorder; neurological disorder; reproductive disorder;
smooth muscle disorder; immunological disorder; gene therapy;
diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.

OS Homo sapiens.

Key	Location/Qualifiers
CDS	174..1334
FT	/*tag= a

AA
PN WO200014251-A2.

PD 16-MAR-2000.

PF 09-SEP-1999; 99WO-US20989.

PR 10-SEP-1998; 98US-0150657.

FR 04-NOV-1998; 9805-0186779.
PR 11-MAY-1999; 9905-0133642.

PA (INCY-) INCYTE PHARM INC

Tang YT, Corley NC, Chee JY

PL Hillman JL, Azimzai Y;
XX

DR WPI; 2000-256996/22.
D-PCR: 28870310

Human transferase proteins useful for preventing, diagnosing and treating cancers and developmental, gastrointestinal, genetic,

PS Claim 9; Page 104-105; 113pp; English.

The present sequence is that of cDNA clone 2617407CBL encoding human transferase RNFSF-11 (see AAY79219), 1 of 15 claimed human transferase proteins of the invention (see AAY79209-23). The clone was isolated from gall bladder cDNA library GBLANOT1. RNFSF-11 is expressed in dermatologic and gastrointestinal tissues, especially those associated with inflammation and cell proliferation. It shows homology to mouse N-acetylglucosamine 6- α -sulfo-transferase. A polynucleotide comprising nucleotides 264-323 or 1272-1331 of the present sequence can be used as a DNA

CC probe. The new human transferases and polynucleotides can be used
CC in the diagnosis, prevention and treatment (including gene therapy
CC and antisense therapy) of cancer, developmental disorders,
CC gastrointestinal disorders, genetic disorders, immunological
CC disorders, neurological disorders, reproductive disorders, and
CC smooth muscle disorders.

Sequence 2065 BP: 466 A; 575 C; 491 G; 533 T; 0 other; XX
SO

Query Match	18.4%;	Score 377;	DB 21;	Length 2065;
Best Local Similarity	63.4%;	Pred. No. 1.5e-53;		
Matches 644;	Conservative	0;	Mismatches 360;	Indels 11

4;

[illegible][illegible]

RESULT 13

AAZ20793
ID AAZ20793 standard; DNA; 1926 BP.

XX
AC
AA730793.XX
DT: 08-DEC-1999 (first entry)

DT 08-DEC-1999 (first entry)

XX
XX
DE Mouse α[vcosv] sulfotransferase-3 coding sequence.

xx Glycosyl sulfotransferase; GSM-3; detection; diagnosis; leukocyte homing;
kw selectin binding interaction; inflammation; lymphocyte homing; mouse;
kw secondary lymph organ; ss.

XX
05
MUS. SP.XX
PN
WC9949018-A1XX
30-SEP-1999XX
PF
26-FEB-1999

XX
PR 20-MAR-1998: 98US-0045284.

PR 12-NOV-1998; 3003 0130511
XX

PA (REFC) UNIV CALIFORNIA
PA (SYNT) SYNTEX USA INC.

PT Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX

XX
DP
WPT: 1999-580442/49

DR p-PSDB; AAY399T9.

Human and murine glycosyl sulfotransferase 3 and related

PT polynucleotides

PS Claim 4; Fig 3; 59pp; English.

This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of the invention. The nucleic acid sequences, probes and primers derived from these, proteins and antibodies are useful in detecting homologues of the sequences, antibodies and methods are useful in the diagnosis and treatment of diseases associated with selectin binding interactions, including conditions associated with or resulting from the homing of leukocytes to sites of inflammation and the normal homing of lymphocytes to secondary lymph organs.

Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;
XX
SN

Query Match	17.3%	Score 354;	DB 20;	Length 1926;
Best Local Similarity	60.9%;			
Matches	614;	Pred. No. 8.9e-50;		
Mismatches	0;			
Conservative	385;			
Indels	9;	Gaps		

QY	341	gtgacatgctgggtgctgctctctggtggcctcgtcctctctctggtggcagctctctc	400
Db	528	gtgatgctgggtgctgctctctggtggcagctctctctctggtggcagctctctc	587
QY	401	aggcagcaccgcgcgtctctcaactgatggagccgcgctggcatgtgtggaccacctg	460
Db	568	gggcagcaccgcgagtgtctctcaactgatggagccgcgctggcatgtgtggagctctc	647
OY	461	tgcagagcagcgcggcgaacgctgcacatggcgcgtgcgcgacatgatgcctctatcttt	520


```

Db      648  accagcagcacagcctggaagctgcacatgctgtgctgggatctctcgtttccgtcttc 707
Qy      521  ttgtcgcacatgacgtgtttgatcctacatg---cacagagccgaacactgtccgcc 577
Db      708  ctgtctgcacatgagcgtgtttgatcctacatgaaccagcccccgaacagtcacgc 767
Qy      578  ttttcaactggcgaacagcgcgtgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 637
Db      768  ctcttccagtgaggaaagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 827
Qy      638  cgaggcaccatcagcaagcagcagctatgcaagaacctgtgcacgcgcgcgcgcgcgcgcgc 697
Db      828  gccacagatcagctcaccacaagcactgcaagctgctcgtcggtcagcagcccttgat 887
Qy      698  ctgcccgcggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 757
Db      888  atggtggaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 947
Qy      758  aacctgcaggtgtctacccgcgtcgcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 817
Db      948  agcctcagccctctaccactactcaggaaccttccctcaacctgcacgtcgtgcac 1007
Qy      818  ctggtgcgcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 877
Db      1008  ctggtcgcagaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1067
Qy      878  cgcgacacgcctgtctgtgggcacaaacgcgaagtgggtggagcgcgcgcgcgcgcgcgcgc 937
Db      1068  gttgacagtcattgtctagggcagcatttgaaacgatacgaaggaagacacgcgcgcgcgc 1127
Qy      938  cgcctgattcgcggaggtgtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 997
Db      1128  tattatgcattgagatcactcgaagaagccaggtggacatagtaagggccatccaaacc 1187
Qy      998  ccgcacaccttcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1057
Db      1188  ctccctgaagctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1247
Qy      1058  ccgctggcagagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1117
Db      1248  ccctggccagacacagactataaattgtgtgggttggtatttttggccacaccc 1307
Qy      1118  gagcctggtatccacacatcaccacgcgggtcggggatcgccagcaagcaatcagagccttc 1177
Db      1308  caaacatgggtttacaatgtcacccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1361
Qy      1178  catacttcgttaggaatgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1237
Db      1362  catactaacccaggaagccctcactcactcactcactcactcactcactcactcactcactcact 1421
Qy      1238  actaagatccctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1297
Db      1422  gaaaaggtttccagcttcaagatgcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1345
Qy      1298  cctgtgtactctcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1345
Db      1482  cadgtcagatctcaacaagacaaggaagcaacctgcccctggatctctcgt 1529

```

RESULT 14

```

AAK91803
ID  AAK91803 standard; cDNA; 877 BP.
XX
AC  AAK91803;
XX
DT  06-NOV-2001 (first entry)
XX
DE  Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX
KW  Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS  Homo sapiens.

```

```

XX      EP1130094-A2.
XX
XX      05-SEP-2001.
XX
XX      07-JUL-2000; 2000EP-0114089.
XX
XX      08-JUL-1999; 99JP-0194486.
XX
XX      11-JAN-2000; 2000JP-0118774.
XX
XX      02-MAY-2000; 2000JP-0183765.
XX
XX      (HELI-) HELIX RES INST.
XX
XX      Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX      WPI; 2001-524255/58.
XX
XX      830 Primers useful for synthesizing full length cDNA clones and their
XX      use in genetic manipulation -
XX
XX      Claim 2; SEQ ID NO 263; 1380pp + sequence listing; English.
XX
XX      The invention relates to primers for synthesising full length cDNA
XX      clones. 830 cDNA molecules encoding a human protein have been
XX      isolated and nucleotide sequences of 5' and 3'-ends of the cDNA
XX      molecules have been determined. Primers for synthesising the full length
XX      cDNA are useful for clarifying the function of the protein encoded by
XX      the cDNA. The full length clones were obtained by construction of full
XX      length enriched cDNA libraries that were synthesised by the oligo-capping
XX      method. The primers enable the production of the full length cDNA easily
XX      without any special methods. The present sequence is the nucleotide
XX      sequence of the 5'-end of a cDNA provided in the invention.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in CD-ROM format directly from EPO.
XX
XX      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

```

```

Query Match      14.6%; Score 297.6; DB 22; Length 877;
Best Local Similarity 71.4%; Pred. No. 1.7e-40;
Matches 406; Conservative 0; Mismatches 160; Indels 3; Gaps 1;

```

```

Qy      335  gatcgtgtgcacgtgctgtcgtcctcgtggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 394
Db      189  gagcgcacgtgcgtgtgtctcctcgtggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 248
Qy      395  ctcttcagccagcaccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 454
Db      249  ctttttggcgcagcccccagatgtttcactgatggagccgcgcgcgcgcgcgcgcgcgcgcgcgc 308
Qy      455  accctgtgcagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 514
Db      309  accctcagcagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 368
Qy      515  atcttttctgcgcacgtgacgtgtttgatcctacatgcgcgcgcgcgcgcgcgcgcgcgcgcgc 571
Db      369  gtctcttctgtgcacatgagcgtctttgatcctacatggaaacctgtcccgcgcgcgcgcgcgc 428
Qy      572  tccgcctttttcaactggcgaacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 631
Db      429  tccagcctcttcagtgagggaacagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 488
Qy      632  ttccccgggcacccatcagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 691
Db      489  atcccacaagatgaatacctccccgggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 548
Qy      692  ttcaagcctgcccgggagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 751
Db      549  tttagtgtgtggaagggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 608
Qy      752  ttctcaacctgcaggtgtctcctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 811

```


662 tttaggtggtggaagcctgcctccacagccacgtggtgctcaaggagtgccg 721
QY 535 ttcttaacctgaggtgctctaccctgtctcagcagcccgctcaacctgcgac 594
Db 722 ttcttaacctgaggtgctctaccctgtctcagcagcccgctcaacctgcgac 781
QY 595 gtgacactgtggtgagcccgccgctgctggtgctcccgagcgccgagcgata 654
Db 782 gtgacactgtggtgagcccgccgctgctggtgctcccgagcgccgagcgata 841
QY 655 ctggcagcgcgaacagcgtgctggtggaacacagcgaagtggtgagccgacact 714
Db 842 ctcatgattgacagcgtgctggtggaacacagcgaagtggtgagccgagac 901
QY 715 cactgctgctgattgctgaggtgctgagcagcagcagcgtgctgagcagc 774
Db 902 caacctactgtgctgaggtgctgagcagcagcagcgtgctgagcagc 961
QY 775 ctcaagcgcgacacttctgctgctgctgctgctgctgctgctgctgctgct 834
Db 962 cagctgctgcaagcctgaggaagcgttactgctgctgctgctgctgctgct 1021
QY 835 egggagcgcgtggaagcctgctgctgctgctgctgctgctgctgctgctgct 894
Db 1022 cgagccctgctggaagcctgctgctgctgctgctgctgctgctgctgctgct 1081
QY 895 cagctgagcctgctgctgctgctgctgctgctgctgctgctgctgctgctg 954
Db 1082 catcttcagcctgctgctgctgctgctgctgctgctgctgctgctgctg 1135
QY 955 gcttcctactgctgctgctgctgctgctgctgctgctgctgctgctgctg 1014
Db 1136 gcttcctacacaaagcaggtgctgctgctgctgctgctgctgctgctgctg 1195
QY 1015 ccttcctactaagcctgctgctgctgctgctgctgctgctgctgctgctg 1074
Db 1196 ccttcctgaaaggttctgctgctgctgctgctgctgctgctgctgctgct 1255
QY 1075 taccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1128
Db 1256 taccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 1309

RESULT 2
US-08-899-514-1
Sequence 1, Application US/08899514
Patent No. 5910581
GENERAL INFORMATION:
APPLICANT: HABUCHI, OSAMI
APPLICANT: FUKUTA, MASAKAZU
TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
TITLE OF INVENTION: FOR THE POLYPEPTIDE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E. ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2156
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Human
TISSUE TYPE: Fetal brain
FEATURE:
NAME/KEY: CDS
LOCATION: 147..1583
IDENTIFICATION METHOD: S
US-08-899-514-1

Query Match 20.1%; Score 235.2; DB 2; Length 2156;
Best Local Similarity 54.0%; Pred. No. 8.e-41;
Matches 589; Conservative 0; Mismatches 468; Indels 33; Gaps 4;

QY 91 gggcctctatcccccagcggcgaggatcgtgtcagcgtgctgtgctgctgtg 150
Db 510 GAGCCGCCAGACGGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 569
QY 151 cgctcggtgctatccttcttggggcagctcttcagcagcagcccccagcgt 210
Db 570 CGCACCGGCTCCCTGCTGGTGGGGGAGTTCTTCAACGAGCGGCGCAACAT 629
QY 211 atgagcccgctggtggtggtggtggtggtggtggtggtggtggtggtg 258
Db 630 TTCGAGCGCGCTGTGGCAGATCGAGCGCAGTGTCTTCGAGCGGGGGGGGG 689
QY 259 gcaacgctgcacatggtggtggtggtggtggtggtggtggtggtggtg 318
Db 690 CGGGGCTCGCCCTGGTGTACCGGAGCTGCTCAAGCAGCTTTCCTGTGCGA 749
QY 319 gtgttgatgctatca-----tgccacagagcgcgaacacgtgcgcctttt 369
Db 750 GTGTGGAGCACTTCATCAGCGCGCTGCCGAGGACACCTGACTCATGTTTCC 809
QY 370 tgggcaacgagccgctgctgctgctgctgctgctgctgctgctgctgctg 429
Db 810 CGGGGCTCCAGCGCTCCCTGTGGGAGGACCCCGCTGTGTAGCCCTTCGTC 869
QY 430 atcagcaagcagcagctgcaagacactgtgacgctggtggtggtggtggt 489
Db 870 TTCGAGAGTACCACTGCAAGAACCGCGCTGCCGCGCCCTCAACGTGAGCT 929
QY 490 gaggcctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 549
Db 930 GAGGCTCCCGCGCAAGAGACATGCGCTCAAGCGGTTGGCGATCCGCGCAT 989
QY 550 gtgctctaccgctgctgagcagcccgctgctgctgctgctgctgctgctg 609
Db 990 TTCCTGAGCGCTGGCGGAGGACCCCGCTGGACCTGGCGCTCATCCAGCTG 1049
QY 610 gaccgctgctgctgctgctgctgctgctgctgctgctgctgctgctgctg 669
Db 1050 GACCCCGGGCGCTGCTGGCTCGCGATGGTGGCTTCCGCGGCAAGTATA 1109
QY 670 ggcctgctggtgagcagcagcagcagcagcagcagcagcagcagcagcag 729
Db 1110 AAGAGTGGCTGGAGCGAGGCGGCGAGCGGCTGAGGAGAGAGAGAGAGAG 1169

QY	730	cgcgaggtgtgcgcgcacacgtgcgcatacgcgcaggccgcacactcaagccgcacacc	789
Db	1170	CGGGGCAACTGTGCAGAG---CATCCGCCCTTCGCGGAGCTGGGGCTCGCGCAGCCCGCC	1226
QY	790	tctctgcgcgcgcctacgcctggtgcgclttcgaggaacctggcgcgggagccgcctggca	849
Db	1227	TGCTTGGGGGGCGCTACATGCTGTGTGGCTACGAGGACGTGGCACCGCGGCGCGCTGCAG	1286
QY	850	gagatccgcgcactctacgcttccacgcctgcacctcacgcacacagctcgaggccttg	909
Db	1287	AAGGCCCGGAGTGTAACCGTTCGCGGCATCCCTTGACCCCGCAGCTGGAGACTGG	1346
QY	910	atocacaacatcacccacggttcgggatcgagcgaagccaatcgaggccttccatactcg	969
Db	1347	ATCCAAAGAACACGACGCGGCCACACGGSCA-----CGGCACTACTCCACG	1397
QY	970	tctaggaaatgcgcgcaagctctcccaggctcgcgccacgcttgcccttcaactagaatc	1029
Db	1398	CAGAAGAACTCTCGGAGCAGTTCCGAGAAGTGGCGTTACGATGCCCCCTTCAAGCTGGCC	1457
QY	1030	ctgcgcgtgcaggagtgtgcgcgcggcgctgcagctgtgtggctacgcgccttgttac	1089
Db	1458	CAGGTGGTCAGGCGCCCCTGGCGCCCTGCCATGCCCTCTCGSGCTACAACATGGCGCGG	1517
QY	1090	tctgcgcgcacacagcgtgacctcaacctggatctgtgtgtgcacagagccacagaccac	1149
Db	1518	GACCGCGCGCCCTCACCACCGCTACGTACGCTGTGTGAGGAGAGGGGCACCTCTGG	1577
QY	1150	ttaagctggg	1159
Db	1578	GTCACTAGG	1587

RESULT 3
US-08-655-878-1
; Sequence 1, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION.

;
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE

CORRESPONDENCE ADDRESS:

STREET.

CITY:

; STATE:

; COUNTRY:

;

COMPUTER RISK

FIELDTON I.
COMPTON I.

OPERATING :

; SOFTWARE

; CURRENT API

APPLICAT:

FILING DATE

CLASSIFICATION:

PRIOR APPL

APPLICANT
FITTING DEPT.

ATTORNEY / ACCOUNTANT

NAME:

; REGISTRAR

REFERENCE

TELECOMMUNI

TELEPHONE

TELEFAX:

; INFORMATION
: SEQUENCE
:

LENGTH.

TYPE: ni

; STRANDEDN


```

Db 1506 cactacctggtggtgcggtacagaggacctggtggagagaccctcgaagacacctacggaga 1565
QY 862 ctctacgctctaccgctgacccctcaccgcccacagctcgagcctggtatccaacatc 921
Db 1566 gftgacgatttgggactgtgtgagcccggaatggagcagtttccctggaacatg 1625
QY 922 accaaggttgcgggagtcggaagccaatcgaggcccttcatactctgctaggaatgcg 981
Db 1626 accagtggctg-----ggctcctctccaagccttctggtatctcagcgaatgcc 1679
QY 982 cgaacgtctccagcctggcgccacagcgttgcccttaactaagatcctgcgctgcag 1041
Db 1680 acgacggccgaatgctgctggcggaccgcccctcaacttccagcagatcaaacagggtggag 1739
QY 1042 gaggtgtgcgcggcgctgcagctgcgtggtgctggtaccgcccctgttactctcggaaccag 1101
Db 1740 gagtttctaccagcccatggcgtcctggtgctatgagcgggtcaacagccctgaggag 1799
QY 1102 cagcgtgacctcaacctggatctgtgtg 1131
Db 1800 gtcaagacctcagcaaacctgcttcgg 1829

RESULT 5
; Sequence 1, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (470)...(1918)
US-09-263-023-1

Query Match 11.3%; Score 133; DB 3; Length 2150;
Best Local Similarity 54.3%; Pred. No. 1.7e-19;
Matches 324; Conservative 0; Mismatches 255; Indels 18; Gaps 2;

QY 109 ggcgcgcgcgagatcgtgtgcacgtgctggtgctgctctctctgctgagcctggcctccttc 168
Db 806 ggaagggaacagcgcagttggtgtgtgtgttcaccacgtggcgcctggcctgctccttc 865
QY 169 ttgggcgcagcttccagcagcaccgcagctcttcttaccctgtgagcgcggtggcat 228
Db 866 ttgcgtgagctcttcacacgaacctgaggtgttcttctctatgagcctgtgtggac 925
QY 229 gttgtgacacacctgtcgcagggcagcgcggaacgtgcacatggccgtgcgcgacctg 288
Db 926 gttgtgcaaaactgtaccgccgggacgcgctttcctcgcagggggcagcgcgggacatg 985
QY 289 atgcgctctatcttttggcagacatggcagcgtgtgttgatgcctacatgcc----- 338

```

```

Db 986 ctgagcgcctctctaacgcctgcgactcttctggtttccagctgtatagcccgaggcagt 1045
QY 339 -----acagagccgaacactgtccgcctttttcaactgggcaacgagccgcgcctgtgc 393
Db 1046 ggggggcgcgaacctcaccactctggcactctttggggcagccactacaagggtggatgc 1105
QY 394 tcgcccgcgcgcctgcagcgcctttcccccaggagccaccatcagcagcagcagcgtatgcag 453
Db 1106 tctcgccactctgctcgtcctaccgcaagaggtcgtcggactggtggacgaccgcgtg 1165
QY 454 ---aacactgtgcacgcgcgcagccattcagcctgcgcggggagcctgcgcctcctcagc 510
Db 1166 tgcataaagtgcacactcagcgcctggcagccttgagagagaggtgcgcaagtaccgc 1225
QY 511 cactgtggtcctcaagagaggtgcctcttcaacctcaggtccttaccgcctgcctcagc 570
Db 1226 acggtggttatcaaggcgcgtgcggctctcgtatgctgtgtgttgccgcctgcttaa 1285
QY 571 gacccgcgcctcaacctgcgcctcgtgcacctgtgcgcacctgtgcgcacctgcgcgcgc 630
Db 1286 gatccagccttgacctcaaggtcatccacctagctacgtgactcctcgtgtgtgcgcgc 1345
QY 631 tcccgggagcgcgcggcccgatactggcagcgcgaacagcagcagcagcagcagcagc 687
Db 1346 tcccgatcgcctcgcgtcagcgcctcctccgggaaagcctacaggtggtgcgaagc 1402

RESULT 6
US-08-804-227C-1
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945

```

```

: TITLE OF INVENTION: OF LUNG CANCER
:
: NUMBER OF SEQUENCES: 86
:
: CORRESPONDENCE ADDRESS:
: ADDRESSSES: SEED AND BERRY LLP
: STREET: 6300 Columbia Center, 701 Fifth Avenue
: CITY: Seattle
: STATE: WA
: COUNTRY: USA
: ZIP: 98104
:
: COMPUTER READABLE FORM:
:

```

```

;
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
;
; CURRENT APPLICATION DATA:
;

```

```

; FILING DATE: 18-MAR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: MAKI, David J. 31,392
; REGISTRATION NUMBER: 210121.456
; REFERENCE/DOCKET NUMBER: 210121.456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-622-4900
; TELEFAX: 206-282-6031
; TELEX:
; INFORMATION FOR SEQ ID NO: 63:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 731 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-040-984-63

Query Match 4.9%; Score 57.8; DB 4; Length 731;
Best Local Similarity 58.4%; Pred. No. 0.00068;
Matches 101; Conservative 0; Mismatches 72; Indels 0; Gaps

QY 515 tgggtctcaaggaggtgcgctttctcaacctgcagtgctctaccgcgtcgcagcacc 574
Db 3 TAGTCATAAAGGGTGTGCGGGTCTTCGACGTGCGGGTCTTGGGGCCACTGCTGCGAGACC 62

QY 575 ccgcgtcaacctgcgcatactgcacctggtgcgcaccaccgcgggcgctgctgcgtccc 634
Db 63 CGGCGCTTGGACCTCAAGGTCATCCACTTGGTGGGTATCCCCCGCGGTGGCGAGTTTAC 122

QY 635 gggagggcggccgcgcatactggcagcgcacacgcgcactgctgctggggcacc 687
Db 122 GGAAGCCGCTGCGCCAGCGGCTCATTCGCTAGACCTACAGGTGGTGGCAGC 175

```

```

? NUMBER OF SEQ ID NOS: 114
? SOFTWARE: Patentin Ver. 2.0
? SEQ ID NO 63
? LENGTH: 731
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: modified_base
?

```



```

; NAME/KEY: modified_base
; LOCATION: (692)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (697)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (698)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (699)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (704)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (705)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (712)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (714)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (717)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (718)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (719)
; OTHER INFORMATION: Where n is a, c, g or t
; NAME/KEY: modified_base
; LOCATION: (723)

```

Query Match
4.7%; Score 54.6; DB 4; Length 4689;

Best Local Similarity 44.2%; Pred. No. 0.0042;
Matches 225; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 582 caactgcgcacatgtgacacgtggtgcgcacccgagcgccgtgctgcgtcccgaggc 641
DB 3408 ccaccacggagccgaacacccctctctctcgcagccgagcggaacaccccgagc 3467
QY 642 ggcgggcccgcatactgacgcgcacacggcgcgcgtgctgggcacacacgcaagtgggt 701
DB 3468 caccacactacccgcgaactcaccgcacatggcgcccggtcaccatcccgctcga 3527
QY 702 ggaggccgacctacactgcgcgtgattgcgaggtgtgcgcagccacacgtgcacatgc 761
DB 3528 cgtgcgcacccacacgcacatgcgcacccctctcgcacgcacatcccgccgagaccc 3587
QY 762 cgaggccgcacacactaacgcgcacacccctctcgcgcgcgcgtcaccgcctggcgctt 821
DB 3588 caccgcgtgctgcacaccccgcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 3647
QY 822 cgaggacctggcgaggagcgctggcagagatccgcgcgcgcgcgcgcgcgcgcgcgcgc 881
DB 3648 ggaggacatgcgcgcacatcctggcggaagacgagcgcgccgaggtctctcgcacac 3707
QY 882 gacctacgcacacagtcgaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 941
DB 3708 gctcgc 3767
QY 942 caagccaatcgc 1001
DB 3768 cagc 3827
QY 1002 gc 1061
DB 3828 gc 3887
QY 1062 gcagctgctgggtacgc 1090
DB 3888 caggc 3916

RESULT 13

US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match 4.7%; Score 54.6; DB 4; Length 36778;
Best Local Similarity 44.2%; Pred. No. 0.0057;
Matches 225; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 582 caactgcgcacatgtgacacgtggtgcgcacccgagcgccgtgctgcgtcccgaggc 641
DB 30398 ccaccacggagccgaacacccctctctcgtcagccgagcgcggaacaccccgagc 30457
QY 642 ggcggccgcgcacatgcgcgcacacggcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 701
DB 30458 caccacactcaccgc 30517

QY 702 ggaggccgcacacactcaccgtgcgcgtgattgcgaggtgtgcgcgacgcacacgtgcacatgc 761
DB 30518 cgtgcgcacccacacgcacatgcgcacccctctcgcacgcacatcccgccgagacgcacct 30577
QY 762 cgaggccgcacacactaacgcgcacacccctctcgcgcgcgcgcgcgcgcgcgcgcgcgc 821
DB 30578 caccgcgtgctgcacaccccgcgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 30637
QY 822 cgaggacctggcgaggagcgctggcagagatccgcgcgcgcgcgcgcgcgcgcgcgcgcgc 881
DB 30638 ggaggacatgcgcgcacatcctggcggaagacgagcgcgccgaggtctctcgcacac 30697
QY 882 gacctacgcacacagtcgaggtgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 941
DB 30698 gctcgc 30757
QY 942 caagccaatcgc 1001
DB 30758 cagc 30817
QY 1002 gc 1061
DB 30818 gc 30877
QY 1062 gcagctgctgggtacgc 1090
DB 30878 catggc 30906

RESULT 14

US-09-320-878-19
; Sequence 19, Application US/09320878A
; Patent No. 6117659
; GENERAL INFORMATION:
; APPLICANT: ASHLEY, Gary
; APPLICANT: BETLACH, Melanie C.
; APPLICANT: BETLACH, Mary C.
; APPLICANT: MCDANIEL, Robert
; APPLICANT: TANG, Li
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
; FILE REFERENCE: 30062202120
; CURRENT APPLICATION NUMBER: US/09/320,878A
; CURRENT FILING DATE: 1999-05-27
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
; EARLIER FILING DATE: 1998-08-28
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
; EARLIER FILING DATE: 1998-05-06
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
; EARLIER FILING DATE: 1997-04-30
; EARLIER APPLICATION NUMBER: 60/119,139
; EARLIER FILING DATE: 1999-02-08
; EARLIER APPLICATION NUMBER: 60/100,880
; EARLIER FILING DATE: 1998-09-22
; EARLIER APPLICATION NUMBER: 60/087,080
; EARLIER FILING DATE: 1998-05-28
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 38506
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-320-878-19

Query Match 4.7%; Score 54.6; DB 3; Length 38506;
Best Local Similarity 44.2%; Pred. No. 0.0057;
Matches 225; Conservative 0; Mismatches 284; Indels 0; Gaps 0;

QY 582 caactgcgcacatgtgacacgtggtgcgcacccgagcgccgtgctgcgtcccgaggc 641
DB 28540 ccaccacggagccgaacacccctctcgtcagccgagcgcggaacaccccgagc 28599

GenCore version 4.5
 Copyright (c) 1993 - 2000 CompuGen Ltd.
 OM protein - protein search, using sw model
 Run on: September 20, 2002, 19:58:31 ; Search time 13.53 seconds
 (without alignments)
 1116.085 Million cell updates/sec
 Title: US-09-593-828-8
 Perfect score: 2051
 Sequence: 1 MWLPRFSKVTVLLAQTT.....LTDLVLPFGDPHFESWASPD 390
 Scoring table:
 BLOSUM62
 Gapop 10.0 , Gapext 0.5
 Searched: 105224 seqs, 38719550 residues
 Total number of hits satisfying chosen parameters: 105224
 Minimum DB seq length: 0
 Maximum DB seq length: 2000000000
 Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries
 Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	540.5	26.4	458	1	C6ST-CHICK
2	105	5.1	359	1	TPSB-CAPEL
3	102.5	5.0	388	1	YM67 MYCTU
4	96	4.7	388	1	ODPT_HUMAN
5	96	4.7	707	1	PAL1 IPOBA
6	95	4.6	885	1	CLPC-ODOSI
7	94.5	4.6	1581	1	VGLP-BEV
8	90.5	4.4	380	1	TPSA-CAPEL
9	89	4.3	371	1	CYB-LATCO
10	88.5	4.3	249	1	NODH-RHTR
11	86.5	4.2	376	1	TPS2-MOUSE
12	86.5	4.2	377	1	TPS2-HUMAN
13	86.5	4.2	1608	1	RRPL-BDV
14	85.5	4.2	221	1	YPRB-CORGL
15	85.5	4.2	389	1	ODPA-FIG
16	85	4.1	1182	1	RTP2-TRVGB
17	84.5	4.1	390	1	ODPA-MOUSE
18	84.5	4.1	390	1	ODPA-RAT
19	84.5	4.1	491	1	C2F1-HUMAN
20	84.5	4.1	1844	1	POLR-TYVVA
21	83.5	4.1	390	1	ODPA-HUMAN
22	83.5	4.1	1844	1	POLR-TYVW
23	83.5	4.1	1844	1	POLR-TYVVC
24	83	4.0	363	1	ODPA-SHMA
25	83	4.0	391	1	ODPT-MOUSE
26	82.5	4.0	1065	1	EMBB-MYCAV
27	82	4.0	746	1	PPK-STRCO
28	81.5	4.0	703	1	PPK-STRCO
29	81.5	4.0	715	1	PERE-HUMAN
30	81.5	4.0	3011	1	POLG-HCV1
31	81	3.9	882	1	Y890-MYCTU
32	81	3.9	972	1	MTRA-YEAST
33	80.5	3.9	704	1	DP3E-RHOCA

34	80.5	3.9	976	1	EPAL_HUMAN
35	80.5	3.9	1942	1	Y054_HUMAN
36	80	3.9	247	1	NOH4-RHME
37	80	3.9	328	1	RPOA-NEIMA
38	80	3.9	392	1	BM15_HUMAN
39	80	3.9	415	1	CXA3-RAT
40	80	3.9	416	1	CXA3-MOUSE
41	80	3.9	933	1	ODOL-ECOLI
42	79.5	3.9	559	1	MASY-MAIZE
43	79	3.9	738	1	PLAK-XENLA
44	78	3.8	462	1	TRPE-THETH
45	77.5	3.8	370	1	TPSI_HUMAN

ALIGNMENTS

RESULT 1	
C6ST-CHICK	
ID AC Q92179;	STANDARD; PRT: 458 AA.
DT 01-NOV-1997 (Rel. 35, Created)	
DT 01-NOV-1997 (Rel. 35, Last sequence update)	
DE Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).	
OS Gallus gallus (Chicken).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC Gallus.	
OX NCBI_TaxID=9031;	
RN [1]	
RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.	
RC STRAIN-WHITE LEGHORN; TISSUE=Embryonic chondrocytes;	
RA MEDLINE=95355490; PubMed=7629189;	
RA Fukuta M., Uchimura K., Nakashima K., Kato M., Kimata K.,	
RA Shinomura T., Habuchi O.;	
RT "Molecular cloning and expression of chick chondrocyte chondroitin 6-	
RT sulfotransferase.";	
RL J. Biol. Chem. 270:18575-18580(1995).	
CC -!- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN	
CC SULFATE.	
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin =	
CC adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.	
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE	
CC (BY SIMILARITY).	
CC -----	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	
CC use by non-profit institutions as long as its content is in no way	
CC modified and this statement is not removed. Usage by and for commercial	
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC or send an email to license@isb-sib.ch).	
CC -----	
CC EMBL; D49915; BAA08655.1; ..	
DR Transferase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.	
DOMAIN 1 23	CYTOSOLSMIC (POTENTIAL).
FT TRANSFER 24 37	SIGNAL-ANCHOR (TYPE II MEMBRANE PROTEIN).
FT DOMAIN 38 458	LOMENAL, CATALYTIC (POTENTIAL).
FT CARBOHYD 62 62	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 73 73	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 95 95	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 236 236	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 399 399	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 443 443	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 458 AA; 52253 MW; C9A3B7D0A5086F0C CRC64;	

Query Match 26.4%; Score 540.5; DB 1; Length 458;
 Best Local Similarity 34.3%; Pred. No. 2.2e-41;
 Matches 124; Conservative 56; Mismatches 141; Indels 31; Gaps 8;
 QY 36 AGGEDRVHVLVLSSWRSGSFLGLFSQHPDVFYLMPEAHV--WTLSQGSAAATHMAV 93

Db 107 AAEPRRHVLLMATRTGSSVGEFFNOQGNFYFLPELWHIERIVTFEPGCANAVGSAL 166
 QY 94 --RDLMSIFLCMDVFDAYM--PQSRNLSEFFNWNATSRALCSPACSAFPRGITSKQD 148
 Db 167 VYRDLQQLLCLDLYLLESTFPAPEHUHTAALFRGSSHSUCEEPVPSLKKVKEKH 226
 QY 149 VCKTLCTROPFLSAREACRSYSHVILKEVRFNQLVLYPLLSDPALNLRIVHVRDPRAV 208
 Db 227 CNRRCGPLNITLAAACRKHMAKTVIRQLFELQPLAEDPRDLRIQLVRDPRAV 286
 QY 209 LRSREAAGPILARDNGIVLTGCKWVE-----ADPHLRIRVCRSHVRIARAATL 259
 Db 287 LVSRMVA-----FSGKYESWKKWAAEAGEAPLQEDVQRIGNCES-IRLSAELGL 335
 QY 260 KPPFFLRGRYRLVRFFEDLAREPLAEIRALYAEFTGLITPOLEAWIHNIHSGSIGKPIEA 319
 Db 336 RQPRWLRGRYMGVRYEDVARAPLRKALEMFRFAGIHPTPOVEEWIRANTQAP---QDSNG 392
 QY 320 FHTSSRNARVNSQAWRHLPFTKILRVOEVCAGALQQLGYRYPVYSADQOQDLTLDLVLPR 379
 Db 393 IYSTQKSSQFQKWRFSIFPKLAQVVDACEPAMRFGYKLASSAQELNRSLSI-LEE 451
 QY 380 GP 381
 Db 452 GP 453

RESULT 2
 TPSS_CABEL STANDARD; PRT; 359 AA.
 ID TPSS_CABEL
 AC Q20351;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative protein-tyrosine sulfotransferase (EC 2.8.2.20)
 DE (Tyrosyl)protein sulfotransferase (TPST).
 GN F42G9.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RA Taich A.;
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CATALYZES THE O-SULFATION OF TYROSINE RESIDUES WITHIN
 CC ACIDIC MOTIFS OF POLYPEPTIDES (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + protein tyrosine =
 CC adenosine 3',5'-bisphosphate + protein tyrosine-O-sulfate.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Golgi membrane (By
 CC similarity).
 CC -1- SIMILARITY: BELONGS TO THE PROTEIN SULFOTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed, usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U00051; AAA91354.1; -;
 DR WormPep; F42G9.8; CE07235.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW Hypothetical protein; Transferase; Transmembrane; Glycoprotein;
 FT SIGNAL-ANCHOR. 1 27 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 28 48 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT TRANSMEM
 FT
 FT LUMENAL, CATALYTIC (POTENTIAL).
 FT DOMAIN 49 359

SQ SEQUENCE 359 AA; 41469 MW; 9FC2F44539B8CC53 CRC64;
 Query Match 5.1%; Score 105; DB 1; Length 359;
 Best Local Similarity 19.8%; Pred. No. 0.049;
 Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
 QY 11 VTVLLAQTCTCLLFI-----ISRPGPSSPAGEDRVH-----VLV 46
 Db 31 IYIFIFCFTICLLIFSSIKCKLQEKLEQISLSKESLIFNEQDARHSRRLLSNLEQITFV 90
 QY 47 LSWRSSSSLGQLFSQHPDV-----FYLMEPAW-HVWTTLSQGSAAATLHMVAVR 94
 Db 91 GGVPSTTTLRALIDRHPVRCGETMLLPSFLTWAGWKNWV---NNSGIT-----141
 QY 95 DLMRSIFLCMDVDYMPQSRNLSAFFNWAISRALCSPACSAFPGTISK-ODVCKTL 153
 Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
 QY 154 CTROPFLSAREACRSYSHVILKEVRFNQLVLYPLLSDPALNLRIVHVRDPRAVLSR- 212
 Db 168 CNKDP-----YTLMLPTIR-----RLYP-----NAKFIIMIRDARAVVHSMI 205
 QY 213 EAAGPILARDNGIVLTGNGKWEADPHLRIRVCRSHVRIAEAAATLKPPLFGRGYLV 272
 Db 206 ERKVPVAGYNTSDISMVQW---NQELRMTFOCNN---APGOCIK-----V 247
 QY 273 RFEDLAREPLAEIRALYAEFTGLITPOLEAWIHNIHSGSIGKPI-----EAFHSS-RNA 327
 Db 248 YYERLIQKPAEELIRITNFDLPFSQOM-----LRHQDLIGDEVLDNQDQFSASQVKN 301
 QY 328 RNVS--QAWRHLPFTKILRVOEVCAGALQQLGY-----RVYS 364
 Db 302 INTXALTSWEDCFSEETWLRKLDV-APFLGILGYDTSISKPDYS 344

RESULT 3
 YM67_MYCTU STANDARD; PRT; 388 AA.
 ID YM67_MYCTU
 AC Q50695;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 46.1 kDa protein RV2267c.
 GN RV2267C OR MT3239 OR MFCY339.43.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Delgic K., Krogan A., McLean J., Moule S., Murphy L.,
 RA Hornsby T., Jagels K., Quail M.A., Rajandream M.A., Rogers J.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischman R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and

laboratory strains.";
Submitted (APR-2001) to the EMBL/GenBank/DBDJ databases.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
or send an email to license@isb-sib.ch).

EMBL; Z77163; CAB00968.1; -;
EMBL; AE007076; AAK46611.1; -;
TIGR; MT2329; -;
TubercuList; Rv2267c; -;
InterPro; IPR00863; Sulfotransferase.
Pfam; PF00685; Sulfotransfer; 1.
Hypothetical protein; Complete proteome.
SEQUENCE 388 AA; 4 6062 MW; 5DED0263275A9B24 CRC64;

[illegible]

RESULT	4
ODPT_HUMAN	
ID	ODPT_HUMAN
PRT;	388 AA.
STANDARD;	
AC	P29803;
DT	01-APR-1993 (Rel. 25, Created)
DT	01-APR-1993 (Rel. 25, Last sequence update)
DE	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Pyruvate dehydrogenase E1 component alpha subunit, testis-specific form, mitochondrial precursor (EC 1.2.4.1) (PDBE1-A type II).
GN	PDHA2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=91065637; PubMed=2249846;
RA	Dahl H.H.M., Brown R.M., Hutchison W.M., Maragos C., Brown G.K.;
RT	"A testis-specific form of the human pyruvate dehydrogenase E1 alpha
RL	subunit is coded for by an intronless gene on chromosome 4.";
RL	Genomics 8:225-232(1990).
CC	-!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
CC	CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
CC	COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
CC	DIHYDROLIPOAMIDE ACETYLTTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
CC	(E3).
CC	-!- CATALYTIC ACTIVITY: Pyruvate + lipoamide = S-

acetyl dihydro lipamide + Co(2).

-!- COFACTOR: THIAMINE PYROPHOSPHATE.

-!- ENZYME REGULATION: E1 ACTIVITY IS REGULATED BY PHOSPHORYLATION (INACTIVATION) AND DEPHOSPHORYLATION (ACTIVATION) OF THE ALPHA SUBUNIT.

-!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.

-!- SUBCELLULAR LOCATION: Mitochondrial matrix.

-!- TISSUE SPECIFICITY: TESTIS, EXPRESSED IN POSTMEIOTIC SPERMATOGONIC CELLS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>) or send an email to license@isb-sib.ch).

EMBL; M66808; AAA60232.1; -

PIR; A37104; DEHUPT.

MIM; 179061; -

InterPro; IPR001017; El_dh.

PFam; PF00676; El_Gehydrog; 1

KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate; Phosphorylation; Mitochondrion; Transit peptide; Multigene family.

FT PHOSPHAT 1 27 MITOCHONDRION (BY SIMILARITY).

FT CHAIN 28 388 PYRUVATE DEHYDROGENASE E1 COMPONENT ALPHA SUBUNIT, TESTIS-SPECIFIC FORM.

FT MOD_RES 230 230 PHOSPHORYLATION (BY SIMILARITY).

FT MOD_RES 291 291 PHOSPHORYLATION (BY SIMILARITY).

FT MOD_RES 298 298 PHOSPHORYLATION (BY SIMILARITY).

FT SEQUENCE 388 AA; 42933 MW; 075B6CFF6DC730C5 CRC64;

	Query Match	4.7%;	Score 96;	DB 1;	Length 388;	
	Best Local Similarity	23.2%;	Pred. No.	0.35;		
	Matches	45;	Conservative	25;	Mismatches	82; Indels
					42;	Gaps
						7;
Qy	82	SQGS-AAFLHMAVRDLMSIFLCLDMVDYDAYMQSRNLSAFFNNWATSRALCSPPACSAATP	140	:	:	:
Dd	198	NQGQTAEAFNWAAWLKPLPCVFIENNLYG-----MGTSTERAASD---DYTK	242	:	:	:
Qy	141	RGTI----SKQDVCKTLCITROPFSLAREACSXYSHVVKEVFENGLVLPILLSDPALNLR	197	:	:	:
Dd	243	KGNFIGPKLVGDMDVLCEVREATAKFYAANCYRSGKGPIIMELQTYRYH--GHNSMSDPGVSYR	300	:	:	:
Qy	198	I-----VHLVRDRPAVLRSEREAAGPIARDNGIVLTNGKWVEADPHLLTEVCVRSHV	251	:	:	:
Dd	301	TREEIQEVRSKRDPPIILQIRM-----VNKLATVTELKEIGAEVRKEID	345	:	:	:
Qy	252	RIAEAATLKPPPFL	265	:	:	:
Dd	346	DAAQFATTDPEPHL	359	:	:	:

RESULT	5
PALI_IPOBA	
ID	PALI_IPOBA STANDARD; PRT; 707 AA.
AC	P4166;
DC	01-JAN-1990 (Rel. 13, Created)
DT	01-JAN-1990 (Rel. 13, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Phenylalanine ammonia-lyase (EC 4.3.1.5).
GN	PAL.
OS	Ipomoea batatas (Sweet potato) (Batatae).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Asteridae; euasterids I; Solanales; Convolvulaceae; Ipomoea.
OX	NCBI_TaxID=4120;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Tanaka Y., MatsuoKa M., Yamanoto N., Ohashi Y., Kano-Murakami Y.,
RA	Ozeki Y.;

RT "Structure and characterization of a cDNA clone for phenylalanine
 RL ammonia-lyase from cut-injured roots of sweet potato.";
 CC Plant Physiol. 90:1403-1407(1989).
 CC !- FUNCTION: THIS IS A KEY ENZYME OF PLANT METABOLISM CATALYZING THE
 CC FIRST REACTION IN THE BIOSYNTHESIS FROM L-PHENYLALANINE OF A WIDE
 CC VARIETY OF NATURAL PRODUCTS BASED ON THE PHENYLPROPANE SKELETON.
 CC !- CATALYTIC ACTIVITY: L-phenylalanine = trans-cinnamate + NH(3).
 CC !- PATHWAY: KEY ENZYME OF PHENYLPROPANOID METABOLISM.
 CC !- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
 CC !- PTM: CONTAINS AN ACTIVE SITE 4-METHYLDENED-IMIDAZOLE-5-ONE (MIO),
 CC WHICH IS FORMED AUTOCATALYTICALLY BY CYCLIZATION AND DEHYDRATION
 CC OF RESIDUES ALA-SER-GLY (BY SIMILARITY).
 CC !- SIMILARITY: BELONGS TO THE PAL / HISTIDASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M29232; AAA33389.1; .
 CC PIR; S29029; S29029.
 CC HSP; P21310; 1B8F.
 CC InterPro; IPR001106; PAL.
 CC Pfam; PF00221; PAL; 2.
 CC PROSITE; PS00488; PAL_HISTIDASE; 1.
 CC KW Lyase; Phenylpropanoid metabolism; Multigene family.
 CC FT SITE 192 194 MODIFIED TO FORM 4-METHYLDENED-IMIDAZOLE-
 CC 5-ONE (BY SIMILARITY).
 CC FT SEQUENCE 707 AA; 77138 MW; 87B9513A91228FBB CRC64;

Query Match 4.7%; Score 96; DB 1; Length 707;
 Best Local Similarity 25.1%; Pred. No. 0.77; 86; Indels 58; Gaps 10;
 Matches 57; Conservative 26; Mismatches 86;

QY 159 FSLAREACRYSYHVVKKE-----VRFNFQVLYPLLS---DPALNLRIVH 200
 DB 131 FGNATESCHILPHSATRAAMLVRINTLQYSGIRFEILEFAITKLNHNITPCPLR-G 188
 QY 201 LVRDPRAVLRREAGPILARDNGIVLTNGKWEADPHILRIEVCVSHVRIEAA--- 257
 DB 189 TITASGDLVPLSYTAGLTGRPNKAVGPNGETLNAEALRL---SRSGRIFRVASPR 244
 QY 258 ---TLKPPPLRCGRYLRVREDIAREPLAPI-RALVA-----TGLRLTPOLE--- 301
 DB 245 KGLPSLMAPPLVLGMAVLEFANVLAVLSEVLSAIFAENVNGKPEFID-HLTHKLKHP 303
 QY 302 -----AWIHNTHSGGIGKPIEAFH-----TSSRNARNVQAW 334
 DB 304 GQIEAAALMEHILDSSVVKAAQKLHEMDPLQKPKQDRYALRTSPQW 350

RESULT 6
 CLPC_ODOSI STANDARD; PRT; 885 AA.
 AC P49574;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE ATP-dependent clp protease ATP-binding subunit clpA homolog.
 GN CLPC.
 OS Odontella sinensis.
 OG Chloroplast.
 OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
 CC Biddulphiophycidae; Eupodiscaceae; Eupodiscaceae; Odontella.
 CX NCBI_taxID=2839;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
 RT "The chloroplast genome of a chlorophyll a+c-containing alga,

RT Odontella sinensis.";
 RL Plant Mol. Biol. Rep. 13:336-342(1995).
 CC !- FUNCTION: MAY INTERACT WITH A CLPP-LIKE PROTEASE INVOLVED IN
 CC DEGRADATION OF DENATURED PROTEINS IN THE CHLOROPLAST.
 CC !- SUBCELLULAR LOCATION: Chloroplast.
 CC !- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z67753; CAA91619.1; .
 CC InterPro; IPR003593; AAA.
 CC InterPro; IPR003959; AAA_subfam.
 CC InterPro; IPR001270; CLP_AB.
 CC InterPro; IPR004176; CLP_N.
 CC Pfam; PF00004; AAA; 1.
 CC Pfam; PF02861; Clp_N; 2.
 CC PRINTS; PR00300; CLPPROTEASEA.
 CC SMART; SM00382; AAA; 1.
 CC PROSITE; PS00870; CLPAB_1; 1.
 CC PROSITE; PS00871; CLPAB_2; 1.
 CC KW Chapterone; ATP-binding; Repeat; Chloroplast.
 CC FT NP_BIND 218 225 ATP (POTENTIAL).
 CC FT NP_BIND 560 567 ATP (POTENTIAL).
 CC FT SEQUENCE 885 AA; 99917 MW; 8F5BE76B5A65380B CRC64;

Query Match 4.6%; Score 95; DB 1; Length 885;
 Best Local Similarity 21.0%; Pred. No. 1.3; 161; Indels 114; Gaps 18;
 Matches 89; Conservative 59; Mismatches 161;

QY 5 RFSSKTVTLVLLIAO-----TCLLLFTII-SRPPGSSPAGGDRVHVVLSSWR 51
 DB 4 KFTGAIKVMILSOEERARRMGHNFGVTEQLLLGIQHRHGIGARALKKQKV---TLKKAR 60
 QY 52 -----SGSFTQLQFSQHPDVFFYLMPEAWHVWTTLSQGSAAFLHMAVRDLMS--IF 101
 DB 61 REIELYIGRGTFGVASEIPFTPRAKRVLEMAVHEGKDLGQNFVGFTEHILLALISESDGVA 120
 QY 102 LCDMDVFDAYMQSRLSAFTNWTATSRALCSPPCASAPRGTSIKQDVCKTLCTQPFTSL 161
 DB 121 MRTLKLGWNTPKGNLLIMVI-----EENGEELRLPTQAEKFL 161
 QY 162 AREACRS-----YSHVVLKEVRFFNLQVLYPLLSDFALNL-RIVHLVROPRAVLRSGE 213
 DB 162 EREKGSSTPTILDEYSENIKEAVDGKL-----DPVIGRDKKEIHEV---IKVLARR 210
 QY 214 AAGPILARDNGIVLTNGKWEADPHILRIEVCVSHVRIEAAATLKPPPLRCGRYLRV 273
 DB 211 KNNPVLLGEPGV-----GKTAVAEGIAQLI-----IAE-----KAPFDLDGML--- 248
 QY 274 PEDLAREPLAETRALYAFGLTLTPOLEAWIHNTHSGGIGKPIEAFH-----SSRNAR 328
 DB 249 ---IMALDLSILACTKYG--EFERIKRIVEVONDSAILIVDEIHTLVGAGAAEGAV 304
 QY 329 NVSQAWRHLPFTKILRVQEVGAGALQLLGYR-----PVYSADQQDRLILD 374
 DB 305 DAANILKALARGKFR---CIGATTIDEYKRIERDPALERRPQFVHVKEFTVGVTIE 359
 QY 375 LVL 377
 DB 360 ILL 362

RESULT 7
 VGLP_BEV STANDARD; PRT; 1591 AA.
 ID VGLP_BEV
 AC P23052;

[illegible]

```

entitles requires a license agreement (See http://www.isb-sib.ch/announce
or send an email to license@isb-sib.ch).
-----
CC BMBL; AF217834; AAF37253.1; -.
CC InterPro: IPR000179; Cyt.b.b6.
DR Pfam: PF00032; cytochrome_b_c; 1.
DR Pfam: PF00033; cytochrome_b_N; 1.
DR PROSITE; PS00193; CYTOCHROME_B_OO; 1.
DR PROSITE; PS00192; CYTOCHROME_B_HEME; FALSE_NEG.
KW Electron transport; Mitochondrion; Respiratory chain; Transmembrane;
Heme.
FT METAL 75 75 IRON 1 (HEME B562 AXIAL LIGAND).
FT METAL 89 89 IRON 2 (HEME B566 AXIAL LIGAND).
FT METAL 174 174 IRON 3 (HEME B562 AXIAL LIGAND).
FT METAL 188 188 IRON 4 (HEME B566 AXIAL LIGAND).
FT SEQUENCE 371 AA; 41947 MW; 36193FF688D6D53C CRC64;
-----
Query Match 4.3%; Score 89; DB 1; Length 371;
Best Local Similarity 23.3%; Pred. No. 1.4;
Matches 53; Conservative 33; Mismatches 76; Indels 66; Gaps 13;
-----
QY 43 HVLVLS-----WRSGSFLLQLSQHPDYVYLMPEPAHVWTLTSGSAAATHLM 91
| | : | : | | | | | | | | | | | | | | : | : | : |
D5 HTLLISNLPEVGSNISTWTFGSMLLSCLQLQTTTGFFL---ALHYTANIMLAFSSVLIH 61
| | : | : | | | | | | | | | | | | | | : | : | : |
QY 92 AVREL-----MRSFLCDMDVFDAYMPQSRNL--SAFFN---WATSRALCSP 133
| | : | : | | | | | | | | | | | | | | : | : | : |
D5 -TRDVPYGMWIMONTHAISASAFFC-----IYIHIARGLYGYSLNKGWVLGTGVALTT 114
| | : | : | | | | | | | | | | | | | | : | : | : |
QY 134 PACSAF-----PRGTIS--KODVCKTCTCTROPF---SLAREACRSYSHVVLKEVRFNLIQ 183
| | : | : | | | | | | | | | | | | | | : | : | : |
D5 LMATAFGYVLPWCMGSFWAATVITNLITAIPLYLSTLTWLGSGFSINDPTLTFRFFALH 174
| | : | : | | | | | | | | | | | | | | : | : | : |
QY 184 VLYPLSDPALNRLVHLVRDPRVLRREAGPILARDNGIVLGTN 230
| | : | : | | | | | | | | | | | | | | : | : | : |
D5 FILPFL---IISLSIIHII-----MLNEGSSNP-----LGTN 204
| | : | : | | | | | | | | | | | | | | : | : | : |
-----
RESULT 10
NODH_RHITR STANDARD; PRT; 249 AA.
ID NODH_RHITR AC P32994;1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Modulation protein H (PC 2.8.2.-).
OS NODH.
GN Rhizobium tropici.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OC NCBI_TaxID=398;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CFN 299;
RX MEDLINE=96303535; PubMed=8755625;
RA Laereans T., Caluwaerts I., Verreth C., Rogel M.A.,
RA Vanderleyden J., Martinez-Romero E.;
RA "Isolation and characterization of Rhizobium tropici Nod factor
sulfolation genes.";
RT Mol. Plant Microbe Interact. 9:492-500(1996).
RL Mol. Plant Microbe Interact. 9:492-500(1996).
RN [2].
RP SEQUENCE FROM N.A.
RC STRAIN=CIAT899;
RX MEDLINE=97002746; PubMed=8850086;
RA Folch-Mallol J.L., Marroqui S., Sousa C., Manyani H., Lopez-Iara I.M.,
RA van der Drift K.M.G.M., Haverkamp J., Quinto C., Gil-Serrano A.,
RA Thomas-Oates J., Spaink H.P., Megias M.;
RA "Characterization of Rhizobium tropici CIAT899 nodulation factors: the
role of nodH and nodPQ genes in their sulfation.";
RT Mol. Plant Microbe Interact. 9:151-163(1996).
RL Mol. Plant Microbe Interact. 9:151-163(1996).
RN [-]
FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.
CC PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO A N-

```


[illegible]

```

RESULT 13
RRPL_BDV
ID RRPL_BDV STANDARD; PRT; 1608 AA.
AC P52639;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE RNA polymerase beta subunit (EC 2.7.7.48) (Large structural protein)
DE (L protein) (P180).
DE L.
OS Borna disease virus (BDV).
OS Viruses; ssRNA negative-strand viruses; Mononegavirales.
OC NCBI_TaxId=12455;
OC [1]
RP SEQUENCE FROM N.A.
RC STRAIN=V;
RC MEDLINE=94240137; PubMed=8183914;
RX Briese T., Schneemann A., Lewis A.J., Park Y.-S., Kim S.,
RA Ludwig H., Lipkin W.I.;
RT "Genomic organization of Borna disease virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4362-4366(1994).
CC -1- FUNCTION: THIS PROTEIN IS PROBABLY A COMPONENT OF THE ACTIVE
CC POLYMERASE.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA}(N).

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch).

CC	EMBL; U04608; AAA02028.1; -
DR	InterPro; IPR01016; Paramyox_RNA_pol.
DR	Pfam; PF00945; Paramyox_RNA_pol; 1.
DR	Transferase; RNA-directed RNA polymerase.
KW	SEQUENCE 1608 AA; 180254 MW; 68F29C42EFE44D57 CRC64;

Query Match	4.2%;	Score 86.5;	DB 1;	Length 1608;
Best Local Similarity	23.7%;	Pred. No. 16;		
Matches 54;	Conservative	31.	Mismatches	92;
			Indels	83;
			Gaps	14;

```
QY 159 FSLAREACRSYSHVVLKEVRFNQLVL-YPLLSDPALNLRIVHLVR-----DPR 207
      :::: -:-:-: -:-:-: -:-:-: -:-:-: -:-:-: -:-:-: -:-:-: -:-:-: -:-:-:
Db 1340 YSLSWSKSRNGSQVLMKMVRUKLCVLYIYPTV-DPAVALDLCHLPALTITLVLGSDPAY 139
```

QY 208 V--LRREAAGPILAR-----DNGIVLGTNGKWEADP-----HLRLIREVCQSHV 251
 Db 1399 YERULEMDLGCAGVSSRVDIPHSLAGTRHGFVSP-----DAGPGVIRLDRLSVCAHP 1453
 QY 252 RTAAEATLKPPPLRGYRLVRFDLAREPLAE-----TRALYAF----- 291
 Db 1454 CUEE---LEFNAYLDS--ELVDISDMCCCLPLATPCKALPRIVRSLSQSFRLALMDNYSFV 1508
 QY 292 -----TGLTLTPQLEAWTHNTHGSSI-----GKPIPAFTSSNAR 328
 Db 1509 MDLIMIRGDIRPHEPDELIVVGQHLGQPVLEVYVGVVRKRPVLRHPWSADLK 1568
 QY 329 NVSQAWRHALPFTKILRVQFVCAGALQLLG 358
 Db 1569 RITVGGAPCAPCSAARLR-DEDCQSL-LVG 1596

RESULT 14
 YPRB_CORGL STANDARD; PRT; 221 AA.
 ID YPRB_CORGL
 AC P46584;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 OS Hypothetical protein in prob 5' region (Fragment).
 OC Corynebacterium glutamicum (Brevibacterium flavum).
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 CC Corynebacterium.
 OX NCBI_TaxID=1718;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 17965 / Melassecola;
 RX MEDLINE=96345604; PubMed=8755867;
 RA Anki S., Serebrjiski I., Reyes O., Leblon G.;
 RT "Mutations in the Corynebacterium glutamicum proline biosynthetic
 pathway: a natural bypass of the proA step";
 RL J. Bacteriol. 178:4412-4419(1996).
 CC -!- SIMILARITY: SOME, TO B.SUBTILIS OBG C-TERMINAL.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U31230; AAC44173.1;
 DR InterPro; IPR000765; GTP1_OBG.
 DR PROSITE; PS00905; GTP1_OBG; PARTIAL.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 221 AA; 24716 MW; A0CA77DB6DF3DB02 CRC64;

Query Match 4.2%; Score 85.5; DB 1; Length 221;
 Best Local Similarity 22.8%; Pred. No. 1.5;
 Matches 53; Conservative 40; Mismatches 88; Indels 51; Gaps 12;
 QY 162 AREACRSYSHVVLKVEFFNQLVYPLL-----SDPALNRIVHLVDPRAVLRSR 212
 Db 14 APEALKSFAEVL--KVRLEFKQFGWPFILISAVARKALDP-LKYKLLIEVQDARK-KRPK 69
 QY 213 EAAGPILARDNGIVLGTNGK-VEADPHLR---LI-----REVCRSVRVTAEAATLKPP 262
 Db 70 EKAESVIKPAVHRVTKGQFQKPEDVEQGGFIITGKPERWIIQDFENDEAV----- 124
 QY 263 PFLGRVRLVRFDLAREPLAEIRALYAFGLTITPOLEAWIHNITHG-----SGICKPI 317
 Db 125 GYLADRLSKLGIEDGLRKAGAHVGANTVIGGISF-----EWEPMITAGDPIILTRGIDV 179

QY 318 EAPHSSRNARNVSOAWRHALPFTKILRVQFVCAGALQLLGIRVYVSADQOR 369
 Db 180 RLQFQTS-----RISAERK-----RASQVRRGLDELIDYGDQASRRR 218
 RESULT 15
 ODPA_PIG STANDARD; PRT; 389 AA.
 ID ODPA_PIG
 AC P29804;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Pyruvate dehydrogenase E1 component alpha subunit, somatic form,
 DE mitochondrial precursor (EC 1.2.4.1) (PDH-E1 type I) (Fragment).
 GN PDHAL.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=90370488; PubMed=2395657;
 RA Sermon K., Demeirleir L., Elpers I., Lissens W., Liebaers I.;
 RT "Characterisation of a cDNA for porcine PDH-E1 alpha and comparison
 RT with the human cDNA";
 RL Nucleic Acids Res. 18:4925-4925(1990).
 CC -!- FUNCTION: THE PYRUVATE DEHYDROGENASE COMPLEX CATALYZES THE OVERALL
 CC CONVERSION OF PYRUVATE TO ACETYL-COA & CO(2). IT CONTAINS MULTIPLE
 CC COPIES OF THREE ENZYMATIC COMPONENTS: PYRUVATE DEHYDROGENASE (E1),
 CC DIHYDROLIPOAMIDE ACETYLTRANSFERASE (E2) & LIPOAMIDE DEHYDROGENASE
 CC (E3).
 CC -!- CATALYTIC ACTIVITY: Pyruvate + lipoamide = S-
 CC acetyldihydrolipoamide + CO(2).
 CC -!- COFACTOR: THIAMINE PYROPHOSPHATE.
 CC -!- ENZYME REGULATION: E1 ACTIVITY IS REGULATED BY PHOSPHORYLATION
 CC (INACTIVATION) AND DEPHOSPHORYLATION (ACTIVATION) OF THE ALPHA
 CC SUBUNIT.
 CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA SUBUNITS.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial matrix.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X52990; CAA37180.1;
 DR PIR; S20813; DEPGA.
 DR InterPro; IPR001017; E1_dh.
 DR Pfam; PF00676; E1_Gehydrog; 1.
 KW Glycolysis; Oxidoreductase; Flavoprotein; Thiamine pyrophosphate;
 KW Phosphorylation; Mitochondrion; Transit peptide.
 FT NON_TER 1
 FT TRANSIT <1 28 MITOCHONDRION (BY SIMILARITY).
 FT CHAIN 29 389 PYRUVATE DEHYDROGENASE E1 COMPONENT
 FT MOD_RES 231 231 ALPHA SUBUNIT, SOMATIC FORM.
 FT MOD_RES 292 292 PHOSPHORYLATION (BY SIMILARITY).
 FT MOD_RES 299 299 PHOSPHORYLATION (BY SIMILARITY).
 SQ SEQUENCE 389 AA; 43121 MW; E9C7DF85389A947 CRC64;

Query Match 4.2%; Score 85.5; DB 1; Length 389;
 Best Local Similarity 17.5%; Pred. No. 3.2;
 Matches 47; Conservative 42; Mismatches 89; Indels 91; Gaps 7;

QY 38 GEDRVHVLVLSWSRSGSFLGOLFQHSQ-----PDVYLMPEFAHWVWTLTSGSAATL 89
 Db 184 GKDEV---CLTLYDGAANQOQIFEAYNMAALWKLPCVFCENNRNYGNGTISVERAASTD 240

```
QY 90 HMAVRDIMSIFLCDDMDVFDAYMPQSRNLSAFTNWATSRALCSPACSAEPRGTISKQDV 149
Db 241 YKRGDFIPGLRVDGMDI-----258
QY 150 CKTLCTRQPFSLAREACRSYSHVVLKEVRFENLOVLYPLLSDPALNLRIVHLVRDPAVL 209
Db 259 ---LCVREATRFAAAYCRSGKGPILMELQTYRYH-----GHMSDQGVSY 300
QY 210 RSREACPIILARDNGIVL---GTNGKWEADPHRLIREVCRSHVRIAFAATLKP--- 261
Db 301 RTREETOEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQAFATADPEPPL 360
QY 262 -----PPF-LRGYRLVRFEDIA 278
Db 361 EELGYHIYCNDPPPEVRGANQWIKFSIS 389
```

Search completed: September 20, 2002, 20:02:46
Job time: 255 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:56:31 ; Search time 21.22 Seconds
(without alignments)
1766.015 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051

Sequence: 1 MWLPRESSKVTVLLLAQT.....LTLDLVLPRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	659.5	32.2	484	2 JE0261	N-acetylglucosamin
2	613.5	29.9	484	2 JC7350	N-acetylglucosamin
3	610	29.7	485	2 JC7331	N-acetylglucosamin
4	540.5	26.4	458	2 A57337	chondroitin 6-sulf
5	107.5	5.2	307	2 F95934	probable enzyme, C
6	105	5.1	359	2 T16350	hypothetical prote
7	102.5	5.0	388	2 G70729	hypothetical prote
8	96	4.7	388	1 DEH0PT	pyruvate dehydroge
9	96	4.7	707	2 S29029	phenylalanine ammo
10	95	4.6	474	2 T38905	probable valine--p
11	95	4.6	885	1 S78246	endopeptidase C1p
12	94.5	4.6	1581	1 VGM03V	peplomer glycoprot
13	94	4.6	582	2 A70841	probable oxalyl-Co
14	93.5	4.6	615	2 C75278	DNA polymerase III
15	93	4.5	523	2 B83629	probable ATP-bindi
16	90.5	4.4	380	2 T42755	tyrosylprotein sul
17	88	4.3	737	2 AD2037	hypothetical prote
18	86.5	4.2	445	2 T05639	hypothetical prote
19	86.5	4.2	577	2 B75595	pyruvate dehydroge
20	86.5	4.2	712	2 T33231	probable long-chai
21	86	4.2	417	2 F71238	hypothetical prote
22	85.5	4.2	221	2 T50665	hypothetical prote
23	85.5	4.2	389	1 DEPGA	hypothetical prote
24	85.5	4.2	425	2 E83023	3-deoxy-D-manno-oc
25	85.5	4.2	571	2 S69210	protein kinase cak
26	85.5	4.2	1254	2 T41262	mutS family DNA mi
27	85	4.1	364	2 A49781	cholesterol dehydr
28	85	4.1	696	2 A11029	hypothetical prote
29	85	4.1	776	2 T31210	membrane protein t

30	85	4.1	1182	2 S14916	hypothetical prote
31	84.5	4.1	334	2 T44163	hypothetical prote
32	84.5	4.1	390	1 DERTP1	pyruvate dehydroge
33	84.5	4.1	390	1 DERTPA	pyruvate dehydroge
34	84.5	4.1	390	2 S23506	pyruvate dehydroge
35	84.5	4.1	491	2 A36036	cytochrome P450 2F
36	84.5	4.1	1844	1 RWRPTM	genome polyprotein
37	84	4.1	552	2 T44954	flagella-related p
38	84	4.1	590	2 T39246	protein phosphotas
39	83.5	4.1	260	2 B95965	probable two-compo
40	83.5	4.1	390	1 DEH0PA	pyruvate dehydroge
41	83.5	4.1	420	2 A53531	oncofetal trophobl
42	83.5	4.1	856	2 F87316	conserved hypothet
43	83.5	4.1	910	2 E89918	2-oxoglutarate deh
44	83.5	4.1	1844	2 S01956	hypothetical prote
45	83.5	4.1	1985	2 S19151	hypothetical prote

ALIGNMENTS

RESULT 1
JE0261
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 07-May-1999
C:Accession: JE0261
R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuo T.
J. Biochem. 124, 670-678, 1998
A>Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
A:Reference number: JE0261; MUID:98391845
A:Accession: JE0261
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DBEJ:AB014679
C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenosine 5-phosphate to the 6-O-sulfate of N-acetylglucosamine.
C:Keywords: sulfotransferase

Query Match	32.2%	Score	659.5	DB 2	Length	484			
Best Local Similarity	40.0%	Pred.	No. 2e-51						
Matches	150	Conservative	67	Mismatches	121	Indels	37	Gaps	12
QY	34	SPAGGEDRVH-VLVSSNWSGSSFLGQLFSQHPDVFYLMPEAWHVWTTLSQGSAAFLHMA	92						
Db	110	APEGVGDKRHWVVTWRSQSSFFGELFNQNPVEFFLYEPVHVWQKLYPGDAYSLOGA	169						
QY	93	VEDLMRSIFLCMDYEDAYMPQ---SRNLS--AFNWNATSRALCSPDPACSPRGTISKQ	147						
Db	170	ARDMLSALRYCDLSVFQLYSPAGSGGRNLTTLGIFGAATNKVVCSSPLCPAYRKEVVGIV	229						
QY	148	D--VKILCTROPFSLAREACRSYSHVVLKEVRFENLOVLYPLLSDDPALNLRIVHLVRDP	205						
Db	230	DDRVCCK-CPPORLARFEECKRYTLVKGVRVEDVAIAPLLRDLPLDLVHLVRDP	288						
QY	206	RAVLRS-----REAGPILARD-----NGIVLTNGKWV--EADPH-LRL	242						
Db	289	RAVASSRIRSRHGLRESLQVVRSDPRAHRMPFEEAAGHKLGAKEGVGGPADYHALGA	348						
QY	243	IEEVCRSHVRTAEATLKPPFRLGRYRLVRFEDLARPLAEIRALYAFGTGLTLPPQLEA	302						
Db	349	MEVICNSMAKTLQTA-LQFPDMLQGHLYLVRYEDLVGEPVKILRRVYFVGLLVSPMEQ	407						
QY	303	WLHNITHGSG-IGKPIEAFHTSSRNARNVSOAHRALPFTKILRVQEVCAQALQILGYEP	361						
Db	408	FALNMTSGSGSSSKP---FVVSARNATQANRALTFTQKQKEEFCYOPMAVLGYER	464						
QY	362	VYSADQQRDLTLDIV	376						
Db	465	VNSPEEVKDLKSTLL	479						

```

RESULT 2
JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse
C:Species: Mus musculus (house mouse)
C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7350
R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462
A:Accession: JC7350
A:Molecule type: mRNA
A:Residues: 1-484 <UCH>
A:Cross-references: DDBJ:AB040710
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.9%; Score 613.5; DB 2; Length 484;
Best Local Similarity 38.8%; Pred. No. 2.9e-47;
Matches 146; Conservative 54; Mismatches 127; Indels 49; Gaps 11;

QY 30 PGSSPAG---GEDRVHVLSSWRSGSFLGOLFQFSDHPDVFYLMFAWHVWTLSSGSA 86
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 85 PGNLSAVGEAVTQEKQHIYVHATWRTGSSFLGELFNQHPDVFYLYEPWHLQALYPGDA 144
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 87 ATLHMVRLMRSIFLCLMDVDFDAY-----MPSQSNLSA--FENWATSRALCSPPA 135
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 145 ESQGLALRMRLSFLRCDFSVRLYAQPGDGERAPDSANLTFAMLEFRWTKVICSPPPL 204
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 136 CSAPPFGT----ISKQDVCKTLCTQPFSL-AREA-CRSYSHVVLKEVRFNQLVLYPLL 189
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 205 CPAAPRARADVLGVEDKACESTC--PPVSLRALAECKKYPVVVKDVLIDGLVLYPLL 262
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 190 SDPALNRIVHLVRDPRVNLRSRAAGPIIARDNGIVLGNK-----WVADP 238
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 263 RDEGLNLKVVQLFRDPRVHNSRLKSRQGLLRISQVLRTRQGDHFRVLLAHGVDPAR 322
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 239 --HLRLREVCRSHVRIAEAA-----TLKPPPLRGYRLVRPFDLAREPLA 283
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 323 GGQARALPSAPRADFFLTLSALEVCEAWLRDLFTFCAPAWLRRYLRXYEDLVWQPQA 382
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 284 EIRALYAFTGLTLPQLEAWTHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKI 343
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 383 QLRLRLRFSGLRLTALDAFAFNMTRGSAVGAD-RPFLHLSADAREAVHVRERLSQDQV 441
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 344 LRQVEVCAGALQLLGY 359
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 442 RQVETACAPAMRLIAY 457
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 3
JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C>Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, N.;
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A:Reference number: JC7350; MUID:20374462
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation
sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in
C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.7%; Score 610; DB 2; Length 486;

```

```

Best Local Similarity 38.8%; Pred. No. 5.9e-47;
Matches 147; Conservative 52; Mismatches 126; Indels 54; Gaps 11;

QY 30 PGSSPAGG---DRVHVLSSWRSGSFLGOLFQFSDHPDVFYLMFAWHVWTLSSGSA 85
| | | | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 86 PSNLSGAVGEAVSEKQHIYVHATWRTGSSFLGELFNQHPDVFYLYEPWHLQALYPGD 145
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 86 AATLHMVRLMRSIFLCLMDVDFDAY-----MQQSRNL--SAFENWATSRALCSPP 134
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 146 AESQGLALRMRLSFLRCDFSVRLYAQPGDPAARADPTANLTFAMLEFRWTKVICSPP 205
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 135 ACSAPPFGTISKQDVCKTLCTQF--PFSL-AREA-CRSYSHVVLKEVRFNQLVLYPLL 190
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 206 LCPGAPRARAEVGLVEDTACERSCPVNAIRALEACKKYPVVVKDVLIDGLVLYPLL 265
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 191 DPALNLRIVHLVRDPR-----VLRSREAAAG-----PIIARDNGIVL 228
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 266 DPGLNLKVVQLFRDPRVHNSRLKSRQGLLRISQVLRTRQGDHFRVLLAHGVGAPRG 325
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 229 TNGKVVADPH-----LRLREVCRSHVRIAEAAATLKPFPFLRGYRLVRPFDLARE 280
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 326 GQSRALPAAPRADFFLTFCALEVCEAWLRDLIFARGA-----PAWLRRYLRXYEDLV 381
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 281 PLAIRALYAFTGLTLPQLEAWTHNTHGSGIGKPIEAFHTSSRNARNVSOAWRHALPF 340
| | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 382 PRAQLRLRFSGLRLTALDAFAFNMTRGSAVGAD-RPFLHLSADAREAVHVRERLSR 440
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 341 TKILREVCAGALQLLGY 359
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 441 EQVQVEACAPAMRLIAY 459
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 4
A57397
chondroitin 6-sulfotransferase (EC 2.8.2.17) precursor - chicken
C:Species: Gallus gallus (chicken)
C>Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000
C:Accession: A57397
R:Fukuta, M.; Uchimura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.; Haba
J. Biol. Chem. 270, 18575-18580, 1995
A:Title: Molecular cloning and expression of chick chondrocyte chondroitin 6-sulfotransferase
A:Reference number: A57397; MUID:95355490
A:Accession: A57397
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-458 <FUK>
A:Cross-references: GB:D49915; NID:g971262; PIDN:BAA08655.1; PID:g971263
C:Superfamily: chondroitin 6-sulfotransferase
C:Keywords: sulfotransferase

Query Match 26.4%; Score 540.5; DB 2; Length 459;
Best Local Similarity 34.3%; Pred. No. 1e-40;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;

QY 36 AGGDRVHVLSSWRSGSFLGOLFQFSDHPDVFYLMFAWHV--WTLSSQSAATLHNAV 93
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 107 AAPEFRVLLMATRTGSSFLGELFNQHPDVFYLYEPWHLQALYPGD 166
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 94 --RLMRSIFLCLMDVDFDAYM---PQSRNLGAFFNWKATSRALCSPPACAPRGITISKQD 148
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 167 YRDVVLQQLLCLDLYLLESTFISPAPEHNTAALFRGSHSLCEPVCVTSKKVFEKXH 226
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 149 VCKTICRQPPSLAREACRSYSHVVLKEVRFNQLVLYPLSDPALNLRIVHLVRDPRV 208
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 227 CKNRRCGFLNITLAAEACERRKQHMALKTVIRLQLEFIQPLAEDPRLDLRIQLVRDPRV 286
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 209 LRSERAGPIIARDNGIVLGNKWE-----ADPHLRILEVCRSHVRIAEATL 259
| | | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 287 LVSRMVA-----FSGYSEYKWKAAEGEAPLQEDVQRLRGNCES-IRLSAEVLG 335
: | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

QY 260 KPPFPLRGYRLVRPFDLAREPLAIRALYAFTGLTLPQLEAWLHNTTHGSGIGKPIEA 319
| : | : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

```
Db 336 ROPRWRGRYMLVRYEDVARAPLKALEMVRFAGIHPTQVEWIRANTQAP--QDSNG 392
QY 320 FHTSSNARNVQAWRHAPFFKILYQVQVACALQQLGTRPVYSADQQRDLIDLVLPR 379
Db 393 ISTQNSSEQFKWRFSPFKLAQVQVQACEPAMFLGYKLASSAQELTNRLSL-LEE 451
QY 380 GP 381
||
Db 452 GP 453

RESULT 5
E95934
A:Description: Probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95934
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: E95934
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-307 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:g15140626; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymb
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,
hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMB21237
A:Genome: plasmid

Query Match 5.2%; Score 107.5; DB 2; Length 307;
Best Local Similarity 20.6%; Pred. No. 0.067;
Matches 78; Conservative 46; Mismatches 131; Indels 123; Gaps 17;

QY 32 PSSPAGGEDRVHVLSSM-RGSSSFLGLFQSHQPDVF---YLMPEAVHVTLSQGSAA 87
||| : : : : : ||| : : : : : ||| : : : : :
Db 2 PSQP-----VRIAYAGRGSSIIIDALQHAHVAGAGHITSLTRVW---RHNEYC 52
||| : : : : : ||| : : : : : ||| : : : : :
QY 88 TLHMAVRLMRSIFLGD-----MDVFDAYMPQSRNLSAFFNWATSRAL 130
||| : : : : : ||| : : : : : ||| : : : : :
Db 53 ACGNAIRD-----CSFSSVVRREWSQDQDPLMEECALQKQFEGLSMMTRLISGMGL 105
||| : : : : : ||| : : : : : ||| : : : : :
QY 131 CSPPACSAFPRGTISKQDVCKTLCITRQPSFLAREACRSYSHVVLKEVRFNQLVYPLIS 190
||| : : : : : ||| : : : : : ||| : : : : :
Db 106 -----GKQFSLYILHTKRLFS-AMQSC-SGRQVIVDSKLPGRMAVAQI- 148
||| : : : : : ||| : : : : : ||| : : : : :
QY 191 DPALMLRVLHVRDPRVLRSEAGPILARDNGIVLGTNGKWEADPHILRLIREVCRSH 250
||| : : : : : ||| : : : : : ||| : : : : :
Db 149 -PGIDMRVHLVRDGRV-----AWSLL-----KGYERDAKSGJQKEI----- 185
||| : : : : : ||| : : : : : ||| : : : : :
QY 251 VRIAPAAVLKPPFFLRGRYR-----LVRFEDLAREPLAETIRALY 289
||| : : : : : ||| : : : : : ||| : : : : :
Db 186 -----KPKSVPTALRWSMNVLAVEYLSRKLGSSEKVMVRVEDPASDPVAVQVQIG 236
||| : : : : : ||| : : : : : ||| : : : : :
QY 290 AFTGLTLFQLEAWTHNTHGSGICKPIEAPHTSSRN-----ARNVQAWRHAPLPT 341
||| : : : : : ||| : : : : : ||| : : : : :
Db 237 TFLFDLS-QVGTSLN-----GEAMGPGHQVAGNRLMNASIALNKDETWRTEMAR 288
||| : : : : : ||| : : : : : ||| : : : : :
QY 342 KILRQVQECAGALQLLGY 359
||| : : : : : ||| : : : : : ||| : : : : :
Db 289 QQVSPQRLGGWMLRRYGY 306
||| : : : : : ||| : : : : : ||| : : : : :
```

```
RESULT 6
T16350
A:Description: hypothetical protein F42G9.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T16350
R:fauch, A.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid F42G9.
A:Reference number: Z18498
A:Accession: T16350
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-359 <TAI>
A:Cross-references: EMBL:U00051; NID:g1216305; PID:g1216308; PIDN:AAA91354.1; CESP:F4
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F42G9.8
A:Introns: 28/2; 50/2; 113/2; 155/3; 220/3; 290/2
C:Superfamily: Caenorhabditis elegans hypothetical protein F42G9.8

Query Match 5.1%; Score 105; DB 2; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.14;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLAQTTCLLLF-----ISRQPSPPAGGEDRVH-----VLV 46
: : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 IYIFIFCFTICLLIFSSIKCKKLEQLSLSKESLIENEQDARHSRLLSNLEQLIFV 90
: : : : : : : : : : : : : : : : : : : : : : : : : :
QY 47 LSSWRSSFLGLFQSHQPDV-----FYLMPEAVHVTLSQGSAAITLMAVR 94
||| : : : : : ||| : : : : : ||| : : : : :
Db 91 GGVSRTGLMRAILDADHPVRCGGTMLPLPSFLTQAGWRNDV---NNSGIT----- 141
||| : : : : : ||| : : : : : ||| : : : : :
QY 95 DLMRSIFLGDMDVFDAYMPQSRNLSAFFNWATSRALCSPPACSAFPRGTISK-QDVCKIL 153
||| : : : : : ||| : : : : : ||| : : : : :
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
||| : : : : : ||| : : : : : ||| : : : : :
QY 154 CTRQPSFLAREACRSYSHVVLKEVRFNQLVYPLISDFALNRLIRVHVRDPRVLRSR- 212
||| : : : : : ||| : : : : : ||| : : : : :
Db 168 CNKDP-----VTALWLPTIR-----RLYP-----NAKFTLMIRDAARVHSMI 205
||| : : : : : ||| : : : : : ||| : : : : :
QY 213 EAAGPILARDNGIVLGTNGKWEADPHILRLIREVCRSHVRIAPAAVLKPPFFLRGRYLV 272
||| : : : : : ||| : : : : : ||| : : : : :
Db 206 ERKVPVAGVNTSDEISMVQW---NOELRKMVTFQCN---APGQCIC-----V 247
||| : : : : : ||| : : : : : ||| : : : : :
QY 273 RFEDLAREPLAETIRALYAFVGLTLTLPQLEAWTHNTHGSGICKPI-----EAPHTSS-RNA 327
||| : : : : : ||| : : : : : ||| : : : : :
Db 248 YYERLQKPAEELIRITNFDLPFSQOM-----LRHQDLIGDEVLDNQEFSSASQVKNS 301
||| : : : : : ||| : : : : : ||| : : : : :
QY 328 RNVS--QAWRHAPLPTKILRQVQECAGALQLLGY-----RPVYS 364
||| : : : : : ||| : : : : : ||| : : : : :
Db 302 INTKALTSMFDCFSERTLRKLDV-APFLGILGYDTSISKPDYS 344
||| : : : : : ||| : : : : : ||| : : : : :

RESULT 7
G70729
A:Description: hypothetical protein Rv2267c - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70729
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A:Reference number: A70500; MUID:98295987
A:Accession: G70729
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-388 <COL>
```

A;Cross-references: GB:Z77163; GB:AL123456; NID:g3261610; PIDN:CA800968.1; PID:e255072;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: RV2267c

Query Match 5.0%; Score 102.5; DB 2; Length 388;
Best Local Similarity 22.7%; Pred. NO. 0.26;
Matches 66; Conservative 37; Mismatches 99; Indels 89; Gaps 15;

Qy	44	VIVLSSNRSGSSFLGOLF--SQH-----PDVFLMEPAWHVWITLQSGSAATLH	90
Db	83	IFIYGHKRTGTHLHELVVDDRHGTGTGVECLAPHHELLTE-----WFAPYVEFLVSKH	137
Qy	91	MAVROLMRSIFLCDMDVFDMQPSRNLSPAFFNNWTSRALCSPACASAPRGCTISKQDVC	150
Db	138	RAMDNMDLSL-----HHQDEDE---FVWC-MQGLPSYLLIAFP-----	172
Qy	151	KTLCTROP-----FSLAREACRSYSHVVLKEVFFNLQVLY-----PLTSDPALNLRI-	198
Db	173	---NRPPQVEEYLDLEQVAPRELE--TWKRTLFERVOQYFRRRTXVILKNPHTSFRIK	226
Qy	199	-----VHLVROPRAVLRSREAAAGPTIARDGNIVL---GNGKWEZADPHILRI	243
Db	227	VLEVFQAKFIHVRDPYVYVSTIHLHKALYRIHGLQOPFDGLDDKKVST--YVDLY	284
Qy	244	REVCRSHVRTAEATLKPPPLFLGRYVLRFVEDLAREPLAEIRALYAFGL	294
Db	285	R-----KLDEGRELVDPST---RFVELAYEDLIDGPEGLRLYQHLGL	324

RESULT 8

DEHPT

pyruvate dehydrogenase (lipoamide) (EC 1.2.4.1) alpha chain precursor, testis-specific

N:Alternate names: pyruvate dehydrogenase complex, E1 component alpha chain

C:Species: Homo sapiens (man)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 15-Oct-1999

C:Accession: A37104

R:Dahl, H.H.M.; Brown, R.M.; Hutchison, W.M.; Maragos, C.; Brown, G.K. Genomics 8, 225-232, 1990

A:Title: A testis-specific form of the human pyruvate dehydrogenase E1alpha subunit is

A:Reference number: A37104; MUID:91065637

A:Accession: A37104

A:Molecule type: mRNA

A:Residues: 1-388 <DAH>

A:Cross-references: GB:M86808; GB:J04769; NID:g190789; PID:AAA60232.1; PID:g190790

C:Genetics:

A:Gene: GDB:PDHA2

A:Cross-references: GDB:1120711; OMIM:179061

A. Map position: 4q22-4q23
 B. Superfamily: pyruvate dehydrogenase (lipoamide) alpha chain; thiamin pyrophosphate-binding site; phosphoprotein; phosphatase
 C. Keywords: flavoprotein; heterotetramer; mitochondrion; oxidoreductase; oxidoreductase; F1-F2/Domain: transit peptide (mitochondrion) #status predicted <IN>
 F.28-388/Product: pyruvate dehydrogenase (lipoamide) alpha chain #status predicted <MA>
 F.183-232/Domain: thiamin pyrophosphate-binding domain homology <TPB>
 F.230/Binding site: phosphate (Ser) (covalent) #status predicted
 F.291/Binding site: phosphate (Ser) (covalent) #status predicted
 F.298/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 4.7%; Score 96; DB 1; Length 388;
Best Local Similarity 23.2%; Pred. No. 0.99;
Matches 45: Conservative 25; Mismatches 82; Indels 42; Gaps 7;

QY	82	SGGS--AATLHNAVRDLMSIFELCDMDVEDAYMQSRNLSAFFNWTASRALCSPACSAFP	140
Db	198	NOGQTAERFNAALWKPCVFCIGNLIG-----MGTSTERAASP---DYK	242
QY	141	RGTI---SKQDVCKTLCTHQPFSLAARACRSYSHVLKVEVFNLQVLYPLLSDDPALNR	177
Db	243	RGNFTPGHKVDGMDVLCVREATKFAANYCRSGKGPILMELQTYRH--GHSMDSDPGVSYR	300
QY	198	I-----VHLVRDPNVLRSREAGPFIARDNGITVLGTNGKWEADPHLRLIEVGRSHV	251

```

Db 301 TREIQVRSKDPITILQDRM-----VNSKLAATVEELKEAGAEVRRKRD 345
QY 252 RIAEAATLKPPPF 265
Db 346 DAAQFATTDPEPH 359

```

RESULT
S29029

phenylalanine ammonia-lyase (EC 4.3.1.5) - sweet potato
C/Species: Ipomoea batatas (sweet potato)
C/Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C/Accession: S29029
R/Tanaka, Y.; Matsuoka, M.; Yamanoto, N.; Ohashi, Y.; Kano-Murakami, Y.; Ozawa, K. (1989) Plant Physiol. 90, 1403-1407, 1989
A/Title: Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase from sweet potato
A/Reference number: S29029
A/Accession: S29029
A/Molecule type: mRNA
A/Residues: 1..707 <F>
A/Cross-references: EMBL:X29232; NID:g168271; PIDN:AAA3389.1; PID:g168272
C/Superfamily: histidine ammonia-lyase
C/Keywords: ammonia-lyase; carbon-nitrogen lyase
F:132-194/Cross-link: 5-imidazolynone (Ala-Gly) #status predicted
F:133/Modified site: dehydroalanylone (Ser) #status predicted

Query Match	4.7%	Score 96;	DB 2;	Length 707;
Best Local Similarity	25.1%	Pred. No. 2.2;		
Matches	57:	Conservative	26:	Mismatches 86;
				Indels 58;
				Gaps 10;

Qy	159	PSLARACRSYSHVVLKE-----VRFNLOVLYPLLS---DPALNURIVH	200
Db	131	FGNATESCHTUPHSATRAAALVRLNTILLOYSIGIRFEILKLLNHNITPCLPAR--G	188
Qy	201	LVRDPRAVLSRSEAGPILARDNGIIVLTGKKVADPHLRLLREVCRSHVIAEAA---	257
		: :	
Db	189	TITASGDLVPLVSYAGLITGRPNKSKAVGPNGETLNAAEALRL-----SRSGRI	244
Qy	258	---TLKPPPLPRGRYRLVRFEDLAREPLAEI-RALYA-----FTGTLITPOLE---	301
Db	245	KGLPSLMAPPYLVGMSKVLFEANVLAVLSVLSAITAEVYNGKPPETD-HLTHKLKHP	303
Qy	302	-----AWIHNITHGSGIGKPIEAFH-----TSSRNARNYSQAW	334
b	304	GCIRAAATMEHTIDGSSVVAAGKLEHMDPLQPKDQRYALRSLSPW	350

RESULT 10

T38905
 C:Probable valine--pyruvate transaminase (EC 2.6.1.66) SPAC56E4.03 [similarity] - first
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 20-Jun-2000
 C:Accession: T38905
 R:Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: Z21813
 A:Accession: T38905
 A:Status: translated from GB/EMBL/DBSJ
 A:Molecule type: DNA
 A:Residues: 1-474 <CON>
 A:Cross-references: EMBL:Z99261; NID:g4038621; PID:g2414656; GSPDB:
 A:Experimental source: strain 972h; Cosmid c56E4

Query Match 4.6%; Score 95; DB 2; Length 474;

QY 102 LCDMDVFOAYMPQSRLNLSAFFNWATSRALCSPPACSAFFRGTTSKQDVCKTLCITROFSSL 161
Db :
Db 121 MRTLDKLGVNDPKLNILMIY-----ENQBEILRPITQAOKFTLL 161
QY 162 ABRAACS-----YSHVLKEVRFFNLQVLYPLSLSDPALNL-RIVHLVRDPRAVLRSE 213
Db :
Db 162 EREKKSSTPTLDYESENISEAVDGKL-----DPVTGRKDEIHEV---IKVLARR 210
QY 214 AAGPILIARDNGVLTGNTGWADPHLRLIREVCRSHVRIAEAAATLKPPFLRGRYELVR 273
Db :
Db 211 KNPVLIIGPGV----GKTAVAEGLAQLI-----IAE-----KAPDFLDGNL---- 248
QY 274 FEDLAREPLAEIRALYAFTGLTPOLEAMHNLTHSGSIGTKGKPIEAFT-----SSNRAR 328
Db :
Db 249 ---LWALDLSILACTYRG-EFFERIKRIVEEVONDSAILVIDEHTLVGAGAEGAV 304
QY 329 NVSQNRHALPFTKILRVQEVCAGALLGY-----PVYSADQRDLTLD 374
Db :
Db 305 DAANILKEPALARGFR-----CIGATTIDERYKIERYDPALERFPQVHVKEFTVGVIIE 359
QY 375 LVI 377
Db :
Db 360 ILL 362

RESULT 12
VGWJVB
peplomer glycoprotein precursor - Berne virus (strain p138/72)
C:Species: Berne virus
C>Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jul-1999
C:Accession: A36759
R:Snijder, E.J.; Den Boon, J.A.; Spaan, W.J.M.; Weiss, M.; Horzinek, M.C.
Virolgy 178, 355-363, 1990
A:Title: Primary structure and post-translational processing of the Berne v
A:Reference number: A36759; MUID:91020973
A:Accession: A36759
A:Molecule type: genomic RNA
A:Residues: 1-1581 <SNT>
A:Cross-references: GB:X52506; NID:g62059; PIDN:CAA36748.1; PID:g62060
C:Genetics:
A:Gene: P
C:Superfamily: Berne virus peplomer glycoprotein
C:Keywords: glycoprotein; transmembrane protein
F:1-19/Domain: signal sequence predicted <SIG>
F:20-1581/Product: peplomer glycoprotein #status predicted <PGP>
F:1547-1572/Domain: transmembrane #status predicted <TMN>
F:25,310,384,494,574,935,969,1267,1297,1327,1385,1389,1428,1431,1438,1483,1

Query Match 4.6%; Score 94.5; DB 1; Length 1581;
Best Local Similarity 17.6%; Pred. No. 8.7;
Matches 81; Conservative 65; Mismatches 135; Indels 179; Gaps

QY 2 WLPRFSKTVTLVLLAQTTCLLLFIISRPGPSAGGEDRHVHLVLSWSRSGSFILGQLF 61
Db :
Db 259 WVVAFQNKATAVLPSELVPAQVKTR-----RLG 289
QY 62 SQHPDVFYLMFAHWHTLTSQGSAAATHMAVRDMRSIFLCDMDVDATMPOQRNSAF 121
Db :
Db 290 VNTPDYFWLVQAYH-----YLSEA-NLSP- 313
QY 122 FKNWATSRALCSPPACSAFFRTGISKQDVCKTLCITROPFLSARAECRSYSHVV-LKEVREF 180
Db :
Db 314 -NYALFSALCN-----SLYQQSATLSLTCGSPFFVAQEC---YNNALYLPDAVET 360
QY 181 NL-----QVLYPL-----LSPALNL-----RIVHLVRDPRAVL---- 209
Db :
Db 361 TLFSTLFSWDYQINPLPNQVLTQNETLOQPAYNIQGQTLSQGRMLNLFKAIVLDFED 420
QY 210 -----RSREAAGP-----ILARDNGVLTGINKWEADPHLRLIREVCRSHVRIAEAAATKP 261
Db :
Db 421 IKFYETNDAPSDDIVVVARQAQLIRYGNEFRICINGIQGV---KGSNNI--ISTLEP 473

QY 262 PPFLRGRLVRFEDLAREPLAEIRALYFT----- 292
 Db 474 HP--AGVIMIAHHHSMWSVAARNSTSFYCVTHSLTFGKLDISTGWFFHTLALPSPVQS 531
 QY 293 -----GLITLTPOLEAMHNIT-----HSGIGKPIEAFHTSSRNARNYS 331
 Db 532 VSMPLLSAAVGVYMPHLEHPIPLTLAQSOYQPSFFNIGNKTI-TLTTQLOAVAYY 590
 QY 332 QAWRHLPFTKILRVQEVCAQALQLGYPVYSADQQRDL 371
 Db 591 TAWFLSVIIVRLPEARRLLTG-VQLVPFTQALLSIKQADL 629
 RESULT 13
 A70841
 probable oxalyl-CoA decarboxylase - Mycobacterium tuberculosis (strain H37Rv)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: A70841
 R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Reiltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987
 A:Accession: A70841
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-582 <COL>
 A:Cross-references: GB:AL021926; GB:AL123456; NID:g3261520; PIDN:CAAL17312.1; PID:g290958
 A:Experimental source: strain H37Rv
 C:Genetics:
 A:Gene: oxca
 C:Superfamily: oxalyl-CoA decarboxylase; thiamin pyrophosphate-binding domain homology

Query Match 4.6%; Score 94; DB 2; Length 582;
 Best Local Similarity 21.4%; Pred. No. 2.5;
 Matches 55; Conservative 44; Mismatches 98; Indels 60; Gaps 11;
 QY 152 TLTCTROPFSIAREACRSYSHVILKEVRFNQLVLYPLSDP-----ALNRIHVILV 202
 Db 2 TTRSASPTVLTDG-----HLVVDALKANDVDITGVVGPITDLARAQAASGRYIGF- 56
 QY 203 RDPRAVLRSEAAGPILARDNGIVLTNGK-WVEADPHLRILREVCRSHVRIAEATLKP 261
 Db 57 RHEASAGNAAAAGFLTARP-GVCLTTSYSGPGLNGLPALANATNCPFMQISGSSS-RP 114
 QY 262 -----PPFLRGRLVRFEDLAREPLAEIRALYFTGLTLPQLEAW 303
 Db 115 MVDLQDGYQDLQDINARPPVKAAYRIGVQDIGRGVARAIR-----TAUSGRPGG 166
 QY 304 ITHNTHSGIGKPIEAFHTSSRNARNYSQAWRHLPFTKILRVQEVCAQALQLGYPVY 363
 Db 167 VYLDIPGDVLQCAVEASAAS-----GATWRPVDPAPELLPAPEADRLDVL----- 213
 QY 364 SADOQRDLTLVLPNG 380
 Db 214 -AAQQRPL---LVLSKG 226
 RESULT 14
 C75278
 DNA polymerase III, tau/gamma subunit - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: C75278
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.N.
 Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896
 A:Accession: C75278
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-615 <WHI>
 A:Cross-references: GB:AE002071; GB:AE000513; NID:g6460218; PIDN:AAF11953.1; PID:g646
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2410
 A:Map position: 1
 C:Superfamily: DNA-directed DNA polymerase III gamma chain

Query Match 4.6%; Score 93.5; DB 2; Length 615;
 Best Local Similarity 23.4%; Pred. No. 3.1;
 Matches 92; Conservative 41; Mismatches 141; Indels 119; Gaps 21;
 QY 4 PRFSKTVTVLLAQTCLLLFIIRPCPSPPAGGEDRVHIVLSSWRSGSFLQQLFSQ 63
 Db 44 PRGVGKTTTARLIAMTA-----NCTGPAPKPGCESECLAV-----RAGS----- 83
 QY 64 HPDVFYLMPEAWHVWTTLSQGSAAATLHMA--VRDLMSIFLCMDM-----VFDAYMPQS 115
 Db 84 HPDVMET-----DAASNNVDVDRDLREKVGLAAMRGKKIYILDEAHMS 129
 QY 116 RNLSAFEFNWATSRALCSPPACSAFPGRTISKQDVCKTLCITROPFSIAREACRSY--SHVV 173
 Db 130 R---AEN-ALLKTLPEEPHEVIFILATPEKIPIILSR-----CQHYRFRRLT 176
 QY 174 LKEV--RFFNLQVLYPLSDP-ALNL--RVHV-LVRDPRAVLRSEAAGPILARDNGIVL 227
 Db 177 SEETAGLAGLVTLEGASADPDALNLGRLADGAMRGESLLERMLAAGIATVTRP----- 231
 QY 228 GTNCKWVEADPHLRILREVCRSHVRIAEATLKPFPFLRGRLVRFEDLAREPLAEIRA 287
 Db 232 -----AVEALGLPGERVGVASALLVGD-AGEATSGAAQ 266
 QY 288 LY--AFTGLTLTPOLEAWIHNIHTEGS-GIGKPIEAFHTSSRNARNYSQAWRHLPFTKIL 344
 Db 267 LYRDGFAARTVVEGLVAFAFGAALHAEILGE-----ECRLEGAEVPRLL 310
 QY 345 RVQEVCAQALQLGYPVYSADQQRDLTLVLL 377
 Db 311 KIQ-----AALDEQAFARFARSADQQ---SLELAL 336
 RESULT 15
 B83629
 Probable ATP-binding component of ABC transporter PA0136 [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: B83629
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337
 A:Accession: B83629
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-523 <STO>
 A:Cross-references: GB:AE004451; GB:AE0045958; PIDN:AAG03526.1; GSPDB:G
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA0136

Query Match 4.5%; Score 93; DB 2; Length 523;
 Best Local Similarity 26.4%; Pred. No. 2.7;
 Matches 37; Conservative 17; Mismatches 60; Indels 26; Gaps 5;

```

QY 13 VLLAOTTCLLLHIIIRPGPSSPAGGEDRVHVLVLSWSRSGSFLGQLFSQHPDVFYLM 72
Db 384 VLALAE-ETIRREFAVKAPGAGAPARS-----LSGGNLQKFIILGREIILQAPRLVAAH 434
QY 73 PAHWVWTTLSQGSAA TLH---MAVRDLMR SIFLCMDVFDAYMPQSENLSAFENWATSRA 129
Db 435 PTWGV---DVGAAALHRA LIALRDAGTALLVVS EDDLEFLLSDR-----IAA 480
QY 130 LCSPPACSAFPRTISKQDV 149
Db 481 LCSGR LCPAVATASAPQKV 500

```

Search completed: September 20, 2002, 19:58:58
Job time: 147 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:56:06 ; Search time 14.8 seconds
(without alignments)
643.647 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRFSKTVVLLAQTT.....LTDLVLRGPDHFSWASPD 390

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A.COMB.pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B.COMB.pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A.COMB.pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B.COMB.pep.*
- 5: /cgn2_6/ptodata/2/iaa/PCUTS.COMB.pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1008	49.1	386	4	US-09-045-284A-2
2	665	32.4	483	3	US-09-263-023-2
3	656.5	32.0	484	3	US-09-263-023-4
4	577.5	28.2	479	2	US-08-899-514-2
5	540.5	26.4	458	2	US-08-655-878-2
6	105	5.1	359	3	US-09-150-133-11
7	105	5.1	359	3	US-09-150-141-11
8	105	5.1	359	4	US-09-374-493-11
9	105	5.1	359	4	US-09-374-824-11
10	105	5.1	359	4	US-09-374-492-11
11	90.5	4.4	380	3	US-09-150-133-9
12	90.5	4.4	380	3	US-09-150-141-9
13	90.5	4.4	380	4	US-09-374-493-9
14	90.5	4.4	380	4	US-09-374-824-9
15	90.5	4.4	380	4	US-09-374-492-9
16	90.5	4.4	384	2	US-08-673-789-9
17	86.5	4.2	376	3	US-09-150-133-7
18	86.5	4.2	376	3	US-09-150-141-7
19	86.5	4.2	376	4	US-09-374-493-7
20	86.5	4.2	376	4	US-09-374-824-7
21	86.5	4.2	376	4	US-09-374-492-7
22	86.5	4.2	377	3	US-09-150-133-5
23	86.5	4.2	377	3	US-09-150-141-5
24	86.5	4.2	377	4	US-09-374-493-5
25	86.5	4.2	377	4	US-09-374-824-5
26	86.5	4.2	377	4	US-09-374-492-5
27	86.5	4.2	566	2	US-08-484-993B-41

28	86.5	4.2	566	2	US-08-484-158B-41	Sequence 41, Appl
29	86.5	4.2	566	2	US-08-484-596A-41	Sequence 41, Appl
30	86.5	4.2	566	2	US-08-480-150A-41	Sequence 41, Appl
31	86.5	4.2	566	3	US-08-458-731-41	Sequence 41, Appl
32	86.5	4.2	566	3	US-08-149-223A-41	Sequence 41, Appl
33	86.5	4.2	1711	3	US-08-369-822C-10	Sequence 10, Appl
34	86.5	4.2	1711	3	US-08-582-776C-10	Sequence 10, Appl
35	86.5	4.2	1711	3	US-08-434-831B-10	Sequence 10, Appl
36	83.5	4.1	390	4	US-09-108-020-36	Sequence 36, Appl
37	83	4.0	15281	2	US-08-471-119A-2	Sequence 2, Appl
38	81.5	4.0	829	4	US-08-444-818-69	Sequence 69, Appl
39	81.5	4.0	1786	4	US-08-444-818-54	Sequence 54, Appl
40	81.5	4.0	2261	4	US-08-444-818-66	Sequence 66, Appl
41	81.5	4.0	2436	4	US-08-444-818-75	Sequence 75, Appl
42	81.5	4.0	2772	4	US-08-444-818-89	Sequence 89, Appl
43	81.5	4.0	2894	2	US-08-466-975A-23	Sequence 23, Appl
44	81.5	4.0	2894	2	US-08-391-671A-23	Sequence 23, Appl
45	81.5	4.0	2894	3	US-08-467-902A-23	Sequence 23, Appl

ALIGNMENTS

RESULT 1
US-09-045-284A-2
; Sequence 2, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107051
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-045-284A-2

Query Match 49.1%; Score 1008; DB 4; Length 386;
Best Local Similarity 52.3%; Pred. No. 1.1e-102;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY	14	LLLAQTTCLLLFIISRP-----	GPSSPAGGEDRVHVLVLSWRSGSFLGQ	59
DB	1	MLLPKMKLLLLFVSOMAILALFHMYSHTSSLSMKAPQPMHVLVLSWRSGSFLGQ	60	
QY	60	LFSQHPDVFYLMFAHWVWTTLSQCSAATLHMVARDLMRSIFLCMDVDVDAVM-PQSEN	118	
DB	61	LFGQHPDVFYLMFAHWVWMTFKQSTAWMLHMVARDLIRAVFLCDMSVFDAYMEPPRRQ	120	
QY	119	SAFTNWTNRALCPDAPCSAPPRTTISKQDVCKTCTRQFSLAREACRSYSHVVLKEVR	178	
DB	121	SSLFQWNSRALCPADACAIIPQDEITIPRAHCRLLCSQPPFVVEKACRSYSHVVLKEVR	180	
QY	179	FFNLQVLYPLISDFAINLRIVHLYRDPRAVLRSEAGPILARDNGIYLGIN-GKWV	237	
DB	181	FFNLQVLYPLKDLNLMHVLVLRDPRAVLRSEAGPILARDNGIYLGIN-GKWV	240	
QY	238	PHLRLEVCRSYHRIAFAATLKP-PPFLGRVLRVLPEDLAREPLAIRALYAFGTGL	296	
DB	241	OPYVVMQVICSQLEIYK--TIQSLPKALQRYLLVYEDLARAPVQTSRMVEFVLEF	298	
QY	297	TPOLEAHNHTGSGTGKPIEAFPTSSRNARNVSWARHALPFTKILRVQEVCAQALQ	356	
DB	299	LPHLQTVHNIIRKGMGD--HAFHTNARDALNVSWARWLSLPYEKYSRLQKACGDAML	356	
QY	357	LGVRPVYSADQQRDLTLDLVLPRGPDHFSWASPD	390	


```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/899,514
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DANIEL E ALTMAN
; REGISTRATION NUMBER: 34,115
; REFERENCE/DOCKET NUMBER: TOXAM21.001AUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 714 760 0404
; TELEFAX: 714 760 9502
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 479
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-899-514-2

```

```

Query Match      28.2%; Score 577.5; DB 2; Length 479;
Best Local Similarity 35.7%; Pred. No. 4.7e-55;
Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps 8;

```

```

QY 32 PSSPAGGEDRVHVLSSWRGSSFLGFSQHPDVFYLMPEAWHWTLS----QGSAA 87
Db 123 PPRVAVGPRRHVILMATRTGSSVFGEFFNQGNIFVFLPELWHIERVSEPGGANAA 182
QY 88 TLHMAVRDLMSIFLCDMDVFDAY---MPQSRNLSAFFNWTATSRALCSPPPACAPFRGTI 144
Db 183 GSALYVRDLKQLFCDLYLVLEHFTTLPEDHLTQFMFRGSSRSICEDPVCPTPVKKVF 242
QY 145 SKQDVCKTLCTQPSLAREACRSYSHVVLKVEVRFNLOVLYPLLSDEPALNLRVHLYRD 204
Db 243 EKYHCKNRCRCPNLVTLAAECRRKEHAKNVRITQLEFQPLAEDPRLDLRVQLVRD 302
QY 205 PRAVLRSEACPIIARDNGIVLG---TNGKWVE-----ADPHLRILREVCRSVHRIA 254
Db 303 PRAVLASRWVA-----FAGKYITWKWKLDDEGQDGLREVEVRLGNCEIS-IRLS 351
QY 255 EAATLKPPPLRGYRLVRREDIAREPLAEIRALYAFGLTILTPQLEAWIINITH----G 310
Db 352 AELGLRQPAWLGRYMLVRYEDVARGPLOKAREMYPFAGIPLTPQVEDWICKNTQAAHDG 411
QY 311 SGIGKPIEAFHTSSRNARNYSQAWRHALPFTTKILRVQEVOCAGALQLLGYRPVYSADQQRD 370
Db 412 SGI-----YSTOKNSSEQFEKWRFSNPFKLAQVQVQDACEPAMRLFGYKLARDAALTN 464
QY 371 LTLDIVLPRG 380
Db 465 RSVSLLEERG 474

```

```

RESULT 5
US-08-655-878-2
; Sequence 2, Application US/08655878
; Patent No. 582713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASAKAZU
; APPLICANT: HASUCHI, OSAMI
; TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:

```

```

; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/655,878
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME:
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:
; TELEFAX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 458
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-655-878-2

```

```

Query Match      26.4%; Score 540.5; DB 2; Length 458;
Best Local Similarity 34.3%; Pred. No. 5.3e-51;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps 8;

```

```

QY 36 AGGEDRVHVLSSWRGSSFLGFLSQHPDVFYLMPEAWHV--WTTLSQGSAAITLHMAV 93
Db 107 AAPEPRRHVILMATRTGSSVFGEFFNQGNIFLPELWHIERVTFEPGGANAVGSAL 166
QY 94 --RDLMSIFLCDMDVFDAY---PQSRNLSAFFNWTATSRALCSPPPACAPFRGTISKOD 148
Db 167 VYRDVLQQLLCLDLYLVLEHFTTLPEDHLTQFMFRGSSRSICEDPVCPTPVKKVF 226
QY 149 VCKILCTQPSLAREACRSYSHVVLKVEVRFNLOVLYPLLSDEPALNLRVHLYRDPRV 208
Db 227 CKNRCCGPLNLTILAAECRRKEHAKNVRITQLEFQPLAEDPRLDLRIQLVDRPRAV 286
QY 209 LRSREAGPIIARDNGIVLGNGKWVE-----ADPHLRILREVCRSVHRIAATL 259
Db 287 LVSRMVA-----FSGKYESWKWAAAGEAPLQEDVORLGNCEIS-IRLSAELGL 335
QY 260 KPPFLRGYRLVRREDIAREPLAEIRALYAFGLTILTPQLEAWIINITHGSGIGKPIEA 319
Db 336 RQPLWLRGYMLVRYEDVARGPLOKAREMYPFAGIHPQVEEWIRANTQAP---QDSNG 392
QY 320 FHTSSRNARNYSQAWRHALPFTTKILRVQEVOCAGALQLLGYRPVYSADQQRDLTDLVLP 379
Db 393 IYSTOKNSSEQFEKWRFSNPFKLAQVQVQDACEPAMRLFGYKLASSAQLTNRSLSL-LEE 451
QY 380 GP 381
Db 452 GP 453

```

```

RESULT 6
US-09-150-133-11
; Sequence 11, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF

```

```
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-11
```

```
Query Match 5.1%; Score 105; DB 3; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
```

```
QY 11 VTVLLAQTTCLLIFI-----ISRPGSPSAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGOLFQSDPDV-----FYLMPEAW-HVWTTLSQGSAAATHMAVR 94
Db 91 GGVPSTGTTLMRAILDADPDVRCGGETHMLLPFLTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPFACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTROPFSLAREACRSYSHVVKVREFNQLVLYPLSDPALNLRIVHLVDRDPAVLRSR- 212
Db 168 CNKDP-----YTALWLPPIR-----RLYP-----NAKFLMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTPNGKWEADPHILRLIREVCRSHVRIAEAAATLKPPFLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NQELRKMTFQCNN---APQCQIK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLPQLEAMWNIHGTSGIGKPI-----EAFHTSS-RNA 327
Db 248 YIERLIQKPAEILRIITNFDLPFSQWQ-----LRHQDLIGDEVLDNQDEFSASQVKN 301
QY 328 RNVS--QAWRHAPFTKILRVQEVACAGALQLGY-----RPVYS 364
Db 302 INTKALTSWEDCFSEETIRKLDV-APFLGILGYDTSISKPDYS 344
```

```
RESULT 7
US-09-150-141-11
; Sequence 11, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150,141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-141-11
```

```
Query Match 5.1%; Score 105; DB 3; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
```

```
QY 11 VTVLLAQTTCLLIFI-----ISRPGSPSAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
```

```
QY 47 LSSWRSGSFLGOLFQSDPDV-----FYLMPEAW-HVWTTLSQGSAAATHMAVR 94
Db 91 GGVPSTGTTLMRAILDADPDVRCGGETHMLLPFLTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPFACSAFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTROPFSLAREACRSYSHVVKVREFNQLVLYPLSDPALNLRIVHLVDRDPAVLRSR- 212
Db 168 CNKDP-----YTALWLPPIR-----RLYP-----NAKFLMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTPNGKWEADPHILRLIREVCRSHVRIAEAAATLKPPFLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NQELRKMTFQCNN---APQCQIK-----V 247
QY 273 RFEDLAREPLAEIRALYAFTGLTLPQLEAMWNIHGTSGIGKPI-----EAFHTSS-RNA 327
Db 248 YIERLIQKPAEILRIITNFDLPFSQWQ-----LRHQDLIGDEVLDNQDEFSASQVKN 301
QY 328 RNVS--QAWRHAPFTKILRVQEVACAGALQLGY-----RPVYS 364
Db 302 INTKALTSWEDCFSEETIRKLDV-APFLGILGYDTSISKPDYS 344
```

```
RESULT 8
US-09-374-493-11
; Sequence 11, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-493-11
```

```
Query Match 5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;
```

```
QY 11 VTVLLAQTTCLLIFI-----ISRPGSPSAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKKLEQLSLSKESLIFNEQDARHSRLLSNLEQLIFV 90
QY 47 LSSWRSGSFLGOLFQSDPDV-----FYLMPEAW-HVWTTLSQGSAAATHMAVR 94
Db 91 GGVPSTGTTLMRAILDADPDVRCGGETHMLLPFLTWQAGWRNDWV---NNSGIT----- 141
QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWATSRALCSPFACSAFPRGTISK-QDVCKTI 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167
QY 154 CTROPFSLAREACRSYSHVVKVREFNQLVLYPLSDPALNLRIVHLVDRDPAVLRSR- 212
Db 168 CNKDP-----YTALWLPPIR-----RLYP-----NAKFLMIRDARAVVHSMI 205
QY 213 EAAGPILARDNGIVLGTPNGKWEADPHILRLIREVCRSHVRIAEAAATLKPPFLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVQW---NQELRKMTFQCNN---APQCQIK-----V 247
```

```

QY 273 RFEDLAREPLAEIRALYAFGTGLTLPQLEAWIHNIHSGIGKPI-----EAFHTSS-RNA 327
Db 248 YIERLIQKPAEEILRITNFDLPFSQOM-----LRHQDLIGDEVDLNDQEFASQVKNS 301

QY 328 RNVS--QAWRHALPFTKILRVQECAGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFCDFSEETLRKLDV-APFLGILGYDTSISKPDYS 344

RESULT 9
US-09-374-824-11
; Sequence 11, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-824-11

```

```

Query Match 5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLAQTTCILLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKLQEKLEQLSKESLIQNEQDARHSRLLSNLEQLIFV 90

QY 47 LSSWSSGSSFLGOLFQSPQDV-----FYLMPEAW-HVWTTLSQGSAAITLHMAVR 94
Db 91 GGVPRSGTILRAILDADHPDVRCCGETMLLPFLTWQAGWRNDWV---NNSGIT-----141

QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWTASRALCSPACSAFFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167

QY 154 CTROPFSLAREACRSYSHVLKEVRFNQLVPLSDPALNLRIVHLVDRPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPITR-----NAKFLMRARAVVHSMI 205

QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILREVCRSHVRIAEAAATLKPPFLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVFQW---NQELRKMIFQCN-----APQCCK------V 247

QY 273 RFEDLAREPLAEIRALYAFGTGLTLPQLEAWIHNIHSGIGKPI-----EAFHTSS-RNA 327
Db 248 YIERLIQKPAEEILRITNFDLPFSQOM-----LRHQDLIGDEVDLNDQEFASQVKNS 301

QY 328 RNVS--QAWRHALPFTKILRVQECAGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFCDFSEETLRKLDV-APFLGILGYDTSISKPDYS 344

```

```

RESULT 10
US-09-374-492-11
; Sequence 11, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:

```

```

; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 11
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-492-11

```

```

Query Match 5.1%; Score 105; DB 4; Length 359;
Best Local Similarity 19.8%; Pred. No. 0.0041;
Matches 80; Conservative 53; Mismatches 131; Indels 140; Gaps 21;

QY 11 VTVLLAQTTCILLFI-----ISRPGSPSPAGGEDRVH-----VLV 46
Db 31 IYIFFCFTICLLIFSSIKCKLQEKLEQLSKESLIQNEQDARHSRLLSNLEQLIFV 90

QY 47 LSSWSSGSSFLGOLFQSPQDV-----FYLMPEAW-HVWTTLSQGSAAITLHMAVR 94
Db 91 GGVPRSGTILRAILDADHPDVRCCGETMLLPFLTWQAGWRNDWV---NNSGIT-----141

QY 95 DLMRSIFLCMDVFDAYMPQSRNLSAFFNWTASRALCSPACSAFFPRGTISK-QDVCKTL 153
Db 142 -----QEVFD-----DAVSAFITEIVAKHSELAPRL 167

QY 154 CTROPFSLAREACRSYSHVLKEVRFNQLVPLSDPALNLRIVHLVDRPRAVLRSR- 212
Db 168 CNKDP-----YTALWLPITR-----NAKFLMRARAVVHSMI 205

QY 213 EAAGPILARDNGIVLGTNGKWEADPHLRILREVCRSHVRIAEAAATLKPPFLRGYRLV 272
Db 206 ERKVPVAGYNTSDEISMVFQW---NQELRKMIFQCN-----APQCCK------V 247

QY 273 RFEDLAREPLAEIRALYAFGTGLTLPQLEAWIHNIHSGIGKPI-----EAFHTSS-RNA 327
Db 248 YIERLIQKPAEEILRITNFDLPFSQOM-----LRHQDLIGDEVDLNDQEFASQVKNS 301

QY 328 RNVS--QAWRHALPFTKILRVQECAGALQLLGY-----RPVYS 364
Db 302 INTKALTSWFCDFSEETLRKLDV-APFLGILGYDTSISKPDYS 344

```

```

RESULT 11
US-09-150-133-9
; Sequence 9, Application US/09150133B
; Patent No. 6060295
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150,133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-133-9

```

```
Query Match          4.4%; Score 90.5; DB 3; Length 380;
Best Local Similarity 19.3%; Pred. No. 0.18;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

QY 51 RSGSFLGQLFSQHPDV-----FYLMEPAW-----HWTTLSQG--SAATLHMAY 93
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 79 RSGTTLRAMLDAHPEVRGCEETRVIPRLNLSQWKSEKWNRLQAGVIGEVINNAI 138
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 94 RDLMSIFLCMDVFDAYMPSQNLNLSAFFNWAISRALCSPACSAFFPGRTISKQDVCKTL 153
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 139 SSFIMEIM-----VGHGDRAPKL 156
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 154 CTROPFSLAREACRSYSHVLKEVFFNLQVLYPLSDPALNLRIVHLVRDPAVLRSE 213
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 157 CNKDPFTMK-----SAVYLKE-----LFP-----NAKYLIMRDGRAVNS-- 192
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 214 AAGPILARD---NGIVLG---TNGKWEADPHLRLIREVCRSHVRIAEAAATIKPPFPLR 266
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 193 ----IISRKVTITGFDLNDFRQCMKNAA---IQIMVDQCES---VGEKNCLK----- 236
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 267 GRYLVRFDLAREPLAEIRALYAFTGLTLPQLEAWIHNTTHGSGIGKPIEAFTSSRN 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 237 -----VYVEQLVHLHPEAQMRRITEFLDI---PWDDKVLH---HEQLIGKDISLNVERSS 285
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

QY 327 ARNVSOA-----WRHALPFTKILRVQEVACAGALQILGYRP 361
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 286 DQVVKPVLNDALIKWGTIPEDVVDVADMSV-APMLRLRIGYDP 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

RESULT 12
US-09-150-141-9
; Sequence 9, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE REFERENCE: 5920.495
; CURRENT APPLICATION NUMBER: US/09/150.141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-150-141-9

Query Match          4.4%; Score 90.5; DB 3; Length 380;
Best Local Similarity 19.3%; Pred. No. 0.18;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

QY 51 RSGSFLGQLFSQHPDV-----FYLMEPAW-----HWTTLSQG--SAATLHMAY 93
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 79 RSGTTLRAMLDAHPEVRGCEETRVIPRLNLSQWKSEKWNRLQAGVIGEVINNAI 138
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 94 RDLMSIFLCMDVFDAYMPSQNLNLSAFFNWAISRALCSPACSAFFPGRTISKQDVCKTL 153
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 139 SSFIMEIM-----VGHGDRAPKL 156
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 154 CTROPFSLAREACRSYSHVLKEVFFNLQVLYPLSDPALNLRIVHLVRDPAVLRSE 213
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 157 CNKDPFTMK-----SAVYLKE-----LFP-----NAKYLIMRDGRAVNS-- 192
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 214 AAGPILARD---NGIVLG---TNGKWEADPHLRLIREVCRSHVRIAEAAATIKPPFPLR 266
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 193 ----IISRKVTITGFDLNDFRQCMKNAA---IQIMVDQCES---VGEKNCLK----- 236
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 267 GRYLVRFDLAREPLAEIRALYAFTGLTLPQLEAWIHNTTHGSGIGKPIEAFTSSRN 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 237 -----VYVEQLVHLHPEAQMRRITEFLDI---PWDDKVLH---HEQLIGKDISLNVERSS 285
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

QY 327 ARNVSOA-----WRHALPFTKILRVQEVACAGALQILGYRP 361
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 286 DQVVKPVLNDALIKWGTIPEDVVDVADMSV-APMLRLRIGYDP 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

RESULT 14
US-09-374-824-9
; Sequence 9, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
```

```
QY 327 ARNVSOA-----WRHALPFTKILRVQEVACAGALQILGYRP 361
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 286 DQVVKPVLNDALIKWGTIPEDVVDVADMSV-APMLRLRIGYDP 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

RESULT 13
US-09-374-493-9
; Sequence 9, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374,493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCI/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 9
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-374-493-9

Query Match          4.4%; Score 90.5; DB 4; Length 380;
Best Local Similarity 19.3%; Pred. No. 0.18;
Matches 66; Conservative 43; Mismatches 108; Indels 125; Gaps 17;

QY 51 RSGSFLGQLFSQHPDV-----FYLMEPAW-----HWTTLSQG--SAATLHMAY 93
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 79 RSGTTLRAMLDAHPEVRGCEETRVIPRLNLSQWKSEKWNRLQAGVIGEVINNAI 138
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 94 RDLMSIFLCMDVFDAYMPSQNLNLSAFFNWAISRALCSPACSAFFPGRTISKQDVCKTL 153
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 139 SSFIMEIM-----VGHGDRAPKL 156
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 154 CTROPFSLAREACRSYSHVLKEVFFNLQVLYPLSDPALNLRIVHLVRDPAVLRSE 213
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 157 CNKDPFTMK-----SAVYLKE-----LFP-----NAKYLIMRDGRAVNS-- 192
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 214 AAGPILARD---NGIVLG---TNGKWEADPHLRLIREVCRSHVRIAEAAATIKPPFPLR 266
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 193 ----IISRKVTITGFDLNDFRQCMKNAA---IQIMVDQCES---VGEKNCLK----- 236
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
QY 267 GRYLVRFDLAREPLAEIRALYAFTGLTLPQLEAWIHNTTHGSGIGKPIEAFTSSRN 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 237 -----VYVEQLVHLHPEAQMRRITEFLDI---PWDDKVLH---HEQLIGKDISLNVERSS 285
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

QY 327 ARNVSOA-----WRHALPFTKILRVQEVACAGALQILGYRP 361
| | | | : : | | | | : : | | | | : : | | | | : : | | | |
Db 286 DQVVKPVLNDALIKWGTIPEDVVDVADMSV-APMLRLRIGYDP 326
| | | | : : | | | | : : | | | | : : | | | | : : | | | |

RESULT 14
US-09-374-824-9
; Sequence 9, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
```

QY 51 RSCSSFLGLFSGHPDV-----FYLMEPAW----HYWTILSQG--SAATLHMAY 93
|||::: : |||:
Db 79 RSGTTTMRAMLDAHPEVRGCETRVPRINLRSQWKSEKEWNLQQAGVGTGEVINAI 138

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 20, 2002, 19:55:26 ; Search time 32.41 seconds
(without alignments)
1336.587 Million cell updates/sec

Title: US-09-593-828-8
Perfect score: 2051
Sequence: 1 MWLPRSSKVTWLLAQTT.....ITLDVLPRGPDHFSWASPD 390

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

- 1: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1980.DAT.*
- 2: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1981.DAT.*
- 3: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1982.DAT.*
- 4: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1983.DAT.*
- 5: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1984.DAT.*
- 6: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1985.DAT.*
- 7: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1986.DAT.*
- 8: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1987.DAT.*
- 9: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1988.DAT.*
- 10: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1989.DAT.*
- 11: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1990.DAT.*
- 12: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1991.DAT.*
- 13: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1992.DAT.*
- 14: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1993.DAT.*
- 15: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1994.DAT.*
- 16: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1995.DAT.*
- 17: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1996.DAT.*
- 18: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1997.DAT.*
- 19: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1998.DAT.*
- 20: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA1999.DAT.*
- 21: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA2000.DAT.*
- 22: /SIDS1/gcgdata/hold-geneseq/geneq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2051	100.0	390	22	AA72639 Human glycosyl sul
2	1865.5	91.0	418	21	AA641947 Human ORF1711
3	1729.5	84.3	395	22	AA72640 Human glycosyl sul
4	1536.5	74.9	395	22	AA72638 Mouse glycosyl sul
5	1013	49.4	386	22	AA93309 Human polypeptide
6	1008	49.1	386	22	AA93918 Human glycosyl sul
7	970	47.3	386	21	AA79219 Human transferase
8	958.5	46.7	388	20	AA739919 Mouse glycosyl sul
9	665	32.4	483	20	AA731656 Mouse N-acetylgluc
10	659.5	32.2	530	22	AA95367 Human protein sequ
11	656.5	32.0	484	20	AA731657 Human N-acetylgluc

12	577.5	28.2	479	19	AAW52863 Glycosaminoglycan
13	540.5	26.4	458	18	AAW06480 Chick chondroitin
14	513.5	25.0	411	19	AAW61100 Keratan sulphate 6
15	184.5	9.0	183	22	AB968582 Drosophila melanog
16	174.5	8.5	363	22	AB64512 Drosophila melanog
17	119.5	5.8	596	22	AA72641 Human glycosyl sul
18	119.5	5.8	1222	22	AA72642 Mouse glycosyl sul
19	112.5	5.5	1207	22	AA72643 Drosophila melanog
20	107.5	5.2	315	22	AB64513 Mouse glycosyl sul
21	105	5.1	359	20	AAV06628 C. elegans tyrosyl
22	105	5.1	359	21	AA84309 A tyrosylprotein s
23	99.5	4.9	928	22	AAE05075 Drosophila melanog
24	99.5	4.9	972	22	AB63015 Drosophila melanog
25	96	4.7	692	22	AB615672 Novel human diagno
26	91.5	4.5	953	22	ABG03420 Novel human diagno
27	90.5	4.4	380	20	AAV06627 C. elegans tyrosyl
28	90.5	4.4	380	21	AA84308 A tyrosylprotein s
29	89	4.3	332	22	ABG06686 Novel human diagno
30	88	4.3	268	22	AAW48176 Thermus thermophil
31	87.5	4.3	664	22	AB63316 Drosophila melanog
32	87	4.2	632	22	AB63316 Drosophila melanog
33	86.5	4.2	330	21	AA834913 Gene 7 human secre
34	86.5	4.2	330	21	AA834914 Human secreted pro
35	86.5	4.2	376	20	AAV06626 Mouse tyrosylprote
36	86.5	4.2	376	21	AA84307 A murine tyrosylpr
37	86.5	4.2	377	21	AA84306 A human tyrosylpro
38	86.5	4.2	377	21	AA84306 A human tyrosylpro
39	86.5	4.2	377	21	AA84306 Amino acid sequenc
40	86.5	4.2	377	22	AA933219 Human polypeptide
41	86.5	4.2	377	22	AA933565 Human polypeptide
42	86.5	4.2	566	15	AA85207 Human zona pelluci
43	86.5	4.2	566	20	AA42479 Human zona pelluci
44	86.5	4.2	566	20	AAW81816 Human ZPB protein
45	86.5	4.2	566	21	AA82214 Human zona pelluci

ALIGNMENTS

RESULT 1
AA72639
ID AA72639 standard; Protein; 390 AA.
XX AC AA72639;
XX AC AA72639;
XX Di 02-MAY-2001 (first entry)
XX XX Human glycosyl sulfotransferase-4alpha (GST-4alpha).
DE DE Human glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
KW KW therapy; selectin binding inhibitor; gene therapy; inflammation;
KW KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
KW KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
KW KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; anaemia;
KW KW Hashimoto's disease; Grave's disease; hypoparathyroidism; psoriasis;
KW KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
KW KW myocarditis; adult respiratory distress syndrome; eczema; rejection;
KW KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
KW KW chromosome 16q23.1.
XX OS Homo sapiens.
XX PN WO200106015-A1.
XX PN 25-JAN-2001.
XX PD 19-JUL-2000; 2000WO-US19741.
XX PF 20-JUL-1999; 99US-0144694.
XX PR 13-JUL-2000; 2000US-0593828.
XX PA (REGC) UNIV CALIFORNIA.

QY 1 MWLPRSSKTVTVLLIAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 Db 28 mwlprrsskvtvlllaqtclllfiisrpspaggedrvhvlvlsrsgsflgql 87
 QY 61 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMAYRDLMSIFLCMDVFDAYM-PQSRNLS 119
 Db 88 fshdpdvflylmpawhvtvltsgsaatlhmavrdlmsiflcmdvfdaymefgprqs 147
 QY 120 AFWNATSRALCSPPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 179
 Db 148 slfwensralcsppacsaafprgtiskqdvcktlctropfslareacrsyshvvlkevrf 207
 QY 180 FNLQVLYPLSDPALNLRVHVRDPRAVLRSPERAGPIARDNGIVLGTNGKWEADPH 239
 Db 208 fnlqvlplsdpalnlrvhvrpdravlrspereagpiardngivlgtngkweadph 267
 QY 240 LRLREVCSHVRIRAEAPLTKPPPLRGYRLVRPDLAREPIAEIRALYATGLTLTPQ 299
 Db 268 lrilrevcrshvrireaeapltkppplrgyrlvrpdlarepiaeiralyatgltltpq 327
 QY 300 LEAWHNTHTSGSIGKPIEAFTTSRRNARNVSOAWRHALPFTKILRVQVOCAGALQLIGY 359
 Db 328 leawhnthtsgsigkpieafttsrrnarnvsoawrhalfptkrlrvqevcagalligyl 387
 QY 360 RPYVSADQQRDLTLVLPRGPDHFSWASPD 390
 Db 388 rpyvsadqqrldtlvlprgpdhfswaspd 418

RESULT 3

AAV72640
 ID AAV72640 standard; Protein; 395 AA.
 AC XX
 AC AAV72640;
 DT 02-MAY-2001 (first entry)
 DE Human glycosyl sulfotransferase-4beta (GST-4beta).

Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocardiitis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection; chromosome 16q23.1.

Homo sapiens.

WO200106015-A1.

25-JAN-2001.

19-JUL-2000; 2000WO-US19741.

20-JUL-1999; 99US-0144694.

13-JUL-2000; 2000US-0593828.

(REGC) UNIV CALIFORNIA.

Rosen SD, Lee JK, Hemmerich S;

WPI; 2001-138471/14.

N-PSDB; AAD02697, AAD02700.

New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for diagnostic and therapeutic agent screening applications - Claim 3; Fig 4B; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocardiitis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.
 CC
 CC Sequence 395 AA;

Query Match 84.3%; Score 1729.5; DB 22; Length 395;
 Best Local Similarity 85.8%; Pred No. 5.1e-181;
 Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

QY 1 MWLPRSSKTVTVLLIAQTCLLLFIISRPSPAGGEDRVHVLVLSWRSGSFLGQL 60
 Db 1 mwlprrsskvtvlllaqtclllfiisrpspaggedrvhvlvlsrsgsflgql 59
 QY 61 FSOHPDVFYLMPEAWHVTTLTSGSAATLHMAYRDLMSIFLCMDVFDAYM-PQSRNLS 120
 Db 60 fshdpdvflylmpawhvtvltsgsaatlhmavrdlmsiflcmdvfdaymptwrnlsd 119
 QY 121 FFWNATSRALCSPPPACSAFPRGTISKQDVCKTLCTROPFSLAREACRSYSHVVLKEVRF 180
 Db 120 lfwavsralscspacsaafprgtiskqdvcktlctropfslareacrsyshvvlkevrf 179
 QY 181 NLOVLYPLSDPALNLRVHVRDPRAVLRSPERAGPIARDNGIVLGTNGKWEADPH 240
 Db 180 nlqvlplsdpalnlrvhvrpdravlrspereagpiardngivlgtngtweadpql 239
 QY 241 RLIREVCSHVRIRAEAPLTKPPPLRGYRLVRPDLAREPIAEIRALYATGLTLTPQ 300
 Db 240 rlvrevcrshvrireaeapltkppplrgyrlvrpdlarepiaeiralyatgltsltqpl 299
 QY 301 EAWHNTHTSGSIGKPIEAFTTSRRNARNVSOAWRHALPFTKILRVQVOCAGALQLIGY 360
 Db 300 eawhnthtsgsigkpieafttsrrnarnvsoawrhalfptkrlrvqevcagalligyl 359
 QY 361 PVSADQQRDLTLVLPRGPDHFSWAS 388
 Db 360 pvsadqqrldtlvlprgpdhfswas 387

RESULT 4

AAV72638
 ID AAV72638 standard; Protein; 395 AA.
 AC XX
 AC AAV72638;
 DT 02-MAY-2001 (first entry)
 DE Mouse glycosyl sulfotransferase-4 (GST-4).

Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive; therapy; selectin binding inhibitor; gene therapy; inflammation; systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes; polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis; glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenitis; Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia; demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis; myocardiitis; adult respiratory distress syndrome; eczema; psoriasis; asthma; hypersensitivity; rheumatic fever; tissue rejection;

Db 121 sslfwensralcsapacdiiipqdeiiprahcrlilscsqpfvevkeacrsysghvvlkev 180
 QY 179 FENLOVPLLLSDPALNLRIVHVRDPRAVLRSEAAAGPILARDNGIVLGTN-GKWEAD 237
 Db 181 ffnlqslvpllkdpnlhnlhivlvdpravrfrsrtkglmidsrvmgqheqklkdd 240
 QY 238 PHLRLIREVCRSHVRIAAATLKP-PPFLRGYRLVRFPEDLAREPLAEIRALYAFITGL 296
 Db 241 qpyyvmqvicsqleiyk--tiqslpkalqeryllvryedlarapvaqtsrmyefvglef 298
 QY 297 TPQLEAWIHNIHSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQVCAGALQL 356
 Db 299 lphlqtwhnitrkgmd--hafhtnrdalnvsqarwslpyekvszrlqacgdaml 356
 QY 357 LGYRPVYSADQORDTLDLVLRGPDHFSWASPD 390
 Db 357 lgyrhvrsegeqrnlldll-----stwtvpe 383

RESULT 6
 AAY39918
 ID AAY39918 standard; Protein; 386 AA.
 AC AAY39918;
 DT 08-DEC-1999 (first entry)
 XX Human glycosyl sulfotransferase-3 protein sequence.
 XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KW selectin binding interaction; inflammation; lymphocyte homing; human;
 KW secondary lymph organ.
 XX Homo sapiens.
 XX WO9949018-A1.
 XX 30-SEP-1999.
 XX 26-FEB-1999; 99WO-US04316.
 XX 20-MAR-1998; 98US-0045284.
 PR 12-NOV-1998; 98US-0190911.
 XX (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTAX USA INC.
 XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
 PI WPI; 1999-580442/49.
 DR N-PSDB; AA220792.
 XX Human and murine glycosyl sulfotransferase 3 and related
 FT polynucleotides
 XX Claim 2; Fig 1; 59pp; English.
 CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 XX Sequence 386 AA;
 SQ

Query Match 49.1%; Score 1008; DB 20; Length 386;
 Best Local Similarity 52.3%; Pred. No. 9, 4e-102;
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 14 LLLAQTTCLLLFIIRP-----GPSSPAGGEDRVHVLVLSWSRSGSFLGQ 59
 Db 1 mlipkkmkllilivsqmailalffhmyshmisslsmkagperhvlvlsrsgsfvqg 60
 QY 60 LFSQHPDVFYLMPEAWHVWTTLSQSSAATTHMAVRDLMSIFILCDMDVFDAYM-POSRL 118
 Db 61 lfqqpddvfylnepawhvwmtfkqstammlhnavrdliravfcdmsvfdaymepgrfq 120
 QY 119 SAFFWATSRALCSPPACSAFPRGTISKQDVCKTLCTROPFSLARACRSYSHVVLKEVR 178
 Db 121 sslfwensralcsapacdiiipqdeiiprahcrlilscsqpfvevkeacrsysghvvlkev 180
 QY 179 FENLOVPLLLSDPALNLRIVHVRDPRAVLRSEAAAGPILARDNGIVLGTN-GKWEAD 237
 Db 181 ffnlqslvpllkdpnlhnlhivlvdpravrfrsrtkglmidsrvmgqheqklkdd 240
 QY 238 PHLRLIREVCRSHVRIAAATLKP-PPFLRGYRLVRFPEDLAREPLAEIRALYAFITGL 296
 Db 241 qpyyvmqvicsqleiyk--tiqslpkalqeryllvryedlarapvaqtsrmyefvglef 298
 QY 297 TPQLEAWIHNIHSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQVCAGALQL 356
 Db 299 lphlqtwhnitrkgmd--hafhtnrdalnvsqarwslpyekvszrlqacgdaml 356
 QY 357 LGYRPVYSADQORDTLDLVLRGPDHFSWASPD 390
 Db 357 lgyrhvrsegeqrnlldll-----stwtvpe 383

RESULT 7
 AAY79219
 ID AAY79219 standard; Protein; 386 AA.
 AC AAY79219;
 DT 19-JUN-2000 (first entry)
 XX Human transferase TRNSFS-11.
 XX Transferase; TRNSFS-11; human; antitumour; cell proliferation;
 KW gastrointestinal disorder; developmental disorder;
 KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle disorder; immunological disorder; inflammation;
 KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH Modified-site 121
 FT /note= "potential O-phosphorylation"
 FT Modified-site 107
 FT /note= "potential O-phosphorylation"
 FT Modified-site 217
 FT /note= "potential O-phosphorylation"
 FT Modified-site 252
 FT /note= "potential O-phosphorylation"
 FT Modified-site 364
 FT /note= "potential O-phosphorylation"
 FT Modified-site 380
 FT /note= "potential O-phosphorylation"
 FT Modified-site 35
 FT /note= "potential O-phosphorylation"
 FT Modified-site 50
 FT /note= "potential O-phosphorylation"
 FT Modified-site 81
 FT /note= "potential O-phosphorylation"
 FT Modified-site 287
 FT /note= "potential O-phosphorylation"
 FT Modified-site 243
 FT /note= "potential O-phosphorylation"
 FT Modified-site 30
 FT /note= "potential N-glycosylation"
 FT Modified-site 308


```

QY 221 RDNGIVLTGKWKVEADPHLRLE-----VCRSHVRIAETA-ATLKPPPELRGRY 269
Db 222 vdnshlv9-----qhletikeedqpyamkiicksqdvkaiqtl--pealqdry 270

QY 270 RLVRFEELAREPLAEIRALYAFGTGLTLPOLAEWIHNITHGSGIGKPIEAFTSSRNARN 329
Db 271 lfryedlvraplaqttrlykfvglflphlqltvyvntvtrkgmqg--hafhtnarnaln 328

QY 330 VSAWNRHALPFTKILRVQECAGALOLLGYRPPVYSADQDRLTDLIV 376
Db 329 vsqawsrwiipyekvsklqdcageamdllylvrsqgqgnslsdlil 375

RESULT 9
AAY31656
ID AAY31656 standard; Protein; 483 AA.
AC AAY31656;
XX
DT 09-NOV-1999 (first entry)
XX
DE Mouse N-acetylglucosamine-6-O-sulfotransferase.
XX
KW N-acetylglucosamine-6-O-sulfotransferase; mouse; GlyCAM-1;
KW L-selectin ligand.
XX
OS Mus musculus.
XX
PN EF943688-A2.
XX
PD 22-SEP-1999.
XX
PF 04-MAR-1999; 99EP-0301530.
XX
PR 24-JUN-1998; 98JP-0177844.
PR 05-MAR-1998; 98JP-0054007.
XX
PA (SEK ) SEIKAGAKU CORP.
XX
PI Habuchi O, Kadomatsu K, Kannagi R, Muramatsu H;
PI Muramatsu T, Uchimura K;
XX
DR WPI; 1999-520337/44.
DR N-PSDB; AAX87820.
XX
PT New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful
PT for synthesis of sugar chains, e.g. GlyCAM-1
XX
PS Claim 1; Page 24-25; 41pp; English.
XX
CC The present sequence represents mouse N-acetylglucosamine-6-O-
CC sulfotransferase, an enzyme capable of transferring a sulfate group
CC from a sulfate group donor to a hydroxyl group at the 6 position of
CC an N-acetylglucosamine residue located at the non-reducing end of
CC an oligosaccharide of formula GlcNAcbeta1-3Galbeta1-4GlcNAc, where
CC GlcNAc = N-acetylglucosamine residue, Gal = galactose residue,
CC beta 1-3 = beta 1-3 glycosidic linkage, and beta 1-4 = beta 1-4
CC glycosidic linkage. The enzyme is useful for the synthesis of
CC sugar chains such as GlyCAM-1, a ligand of L-selectin that is
CC involved in homing of lymphocytes and rolling of leukocytes
CC occurring at the early stage of inflammation. DNA encoding the
CC enzyme (see AAX87820) is expected to be used for the large-scale
CC production of N-acetylglucosamine-6-O-sulfotransferase, or
CC artificial synthesis of GlyCAM-1 using transformants which harbour
CC the DNA.
XX
SQ Sequence 483 AA;

```

Query Match 32.4%; Score 665; DB 20; Length 483;
 Best Local Similarity 40.4%; Pred. No. 6.4e-64;
 Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;

```

QY 37 GGDRVHVLYLSSWRSGSTGLGOLFSDHPVFXLMPEAWHWTTLSOGSAATLHWAVRDL 96
Db 113 ggdkrqlvftwrsstsfgefndnpevfilyepvhwvqklypgdavslggaardm 172

QY 97 MRSFTLCDMDVFDAYMPQ---SRNLS--AFFNWATSRALCSPACSAFPRGTISKQD--V 149
Db 173 lsalycrdlsfvglspagsggnlttlgfgaatnkvcvssplcpayrkevqldvdrv 232

QY 150 CKTLCTRPFSLARACRSYSHVVVKVRFNQLVYPLLSDPALNLRIVHLVXDPRAVL 209
Db 233 ckk-cppgrlarfeecerkvrtvkvgrvfdvavlaplikdpaldikvhlvdrprava 291

QY 210 RS-----REAGPILARD-----NGVILGT--NGKWVEADPH-LRLIREV 246
Db 292 ssrlsrshglireslqvrrsrdprahmpfleaaaghkigakkeggpdpadyhalgamevi 351

QY 247 CRSHVRIAETAATLKPPPELRGRVLRVFEELAREPLAEIRALYAFGTGLTLPOLAEWIHN 306
Db 352 cnsmaaktlqta-lqppdwlqghylvryedlvqdpvktlrrvydfvglvyspemeqfaln 410

QY 307 ITHGSG-IGKPIEAFTSSRNARNVSOAHRHALPFTKILRVQECAGALOLLGTRPVYSA 365
Db 411 mtsgsgssskp---fvvsarnatqanawrtaltfgikgveefcyqpmavlgvyrnsp 467

QY 366 DOORDLTDLIV 376
Db 468 evvdkslktil 478

RESULT 10
AAB95367
ID AAB95367 standard; Protein; 530 AA.
XX
AC AAB95367;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17679.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17679; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

```

CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 530 AA;

Query Match 32.2%; Score 659.5; DB 22; Length 530;
Best Local Similarity 39.1%; Pred. No. 3e-63;
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

QY 31 GSSPAG-----GEDRVHVLSSWRSGSFLGQFSDHPDVFYLMPEAHWVITL 81
DB 145 gvaappgngtrgtggvgdkrgivvfttrsgssffgelfnqnpvfflyepwhwqkl 204
QY 82 SGGSAATHMAYRDLMSRSLFCDMDVFDAYMPQ---SRNLS--AFFNWTATRALCSPAC 136
DB 205 ypgdavsllggaardmsalrycdlsvqlspagsggrnlttligfaatkncvcsppic 264
QY 137 SAPFRGTISKQD--VCKTLCTROPFSLAREACRSYSHVVLKEVRFNLOVLPULLSDPAL 194
DB 265 payrkevglvdrvcck-cppgrlarfeecckrkytlvikgrvfdvavlapllrdpal 323
QY 195 NLRIVHLDPRAVILRS-----REAGPILARD-----NGIVLTNGKWV 234
DB 324 dikvhlvdrpravassrirsrhglreslqvrsrdprahmpfleagghkgakkev 383
QY 235 --EADPH-LRLIREVCRSHVRIAEATLKPPPLRGVRLVREFDLAREPLAEIATYAF 291
DB 384 gpaadyhalgamevicsnsmaklqta-lqppdwlghylvrvydvdpvktlrvydf 442
QY 292 TGLTLPOLAEWIHNTHGSG-IGKPIEFATHTSSRNARNVSOAWRHALPFTKILRVQVC 350
DB 443 vglvspemeqfalmtsgssskp---fvvsarnatqaanawrtaltfqgikvvefc 499
QY 351 AGALQLLGYRPVYSADQQRDLTLDLV 376
DB 500 yqpmavlgvrvnspeevkdsktll 525

RESULT 11

AAAY31657
ID AAAY31657 standard; Protein; 484 AA.

XX
AC AAAY31657;

XX
DT 09-NOV-1999 (first entry)

XX Human N-acetylglucosamine-6-O-sulfotransferase.

XX N-acetylglucosamine-6-O-sulfotransferase; human; GlyCAM-1;
XX L-selectin ligand.

OS Homo sapiens.

XX
FN EP943688-A2.

XX

PD 22-SEP-1999.
XX
PF 04-MAR-1999; 99EP-0301530.
XX
PR 24-JUN-1998; 98JP-0177844.
PR 05-MAR-1998; 98JP-0054007.
XX
PA (SEGK) SETKAGAKU CORP.
XX
PI Habuchi O, Kadomatsu K, Kannagi R, Muramatsu H;
PI Muramatsu T, Uchimura K;
XX
DR WPI; 1999-520337/44.
XX N-PSDB; AAH87821.
PT New polypeptides of N-acetylglucosamine-6-O-sulfotransferase, useful
for synthesis of sugar chains, e.g. GlyCAM-1
XX
PS Claim 2; Page 28-30; 4lpp; English.

CC The present sequence represents human N-acetylglucosamine-6-O-
CC sulfotransferase, an enzyme capable of transferring a sulfate group
CC from a sulfate group donor to a hydroxyl group at the 6 position of
CC an N-acetylglucosamine residue located at the non-reducing end of
CC an oligosaccharide of formula GlcNAc β 1-3Gal β 1-4GlcNAc, where
CC GlcNAc = N-acetylglucosamine residue, Gal = galactose residue,
CC beta 1-3 = beta 1-3 glycosidic linkage, and beta 1-4 = beta 1-4
CC glycosidic linkage. The enzyme is useful for the synthesis of
CC sugar chains such as GlyCAM-1, a ligand of L-selectin that is
CC involved in homing of lymphocytes and rolling of leukocytes
CC occurring at the early stage of inflammation. DNA encoding the
CC enzyme (see AAH87821) is expected to be used for the large-scale
CC production of N-acetylglucosamine-6-O-sulfotransferase, or
CC artificial synthesis of GlyCAM-1 using transformants which harbour
CC the DNA.

XX Sequence 484 AA;

Query Match 32.0%; Score 656.5; DB 20; Length 484;
Best Local Similarity 39.7%; Pred. No. 5.5e-63;
Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGGEDRVH-VLVLSSWRSGSFLGQFSDHPDVFYLMPEAHWVITLSSQSAATHMA 92
DB 110 apegvgdkthwvfttrsgssffgelfnqnpvfflyepwhwqklpygdavslqga 169
QY 93 VRDLMSRSLFCDMDVFDAYMPQ---SRNLS--AFFNWTATRALCSPACAPRGVTSKQ 147
DB 170 ardmalsalrycdlsvqlspagsggrnlttligfaatkncvcsppicpayrkevglv 229
QY 148 D--VCKTLCTROPFSLAREACRSYSHVVLKEVRFNLOVLPULLSDPALNLRIVHLDPR 205
DB 230 ddrvcck-cppgrlarfeecckrkytlvikgrvfdvavlapllrdpalldkvlhldpr 288
QY 206 RAVLSR-----REAGPILARD-----NGIVLTNGKWV--EADPH-LRL 242
DB 289 ravassrirsrhglreslqvrsrdprahmpfleagghkgakkevvggadyhalga 348
QY 243 TREVCRSHVRIAEATLKPPPLRGVRLVREFDLAREPLAEIATYAFGLTLPQLEA 302
DB 349 mevicsnsmaklqta-lqppdwlghylvrvydvdpvktlrvydfvlgvlspevmeq 407
QY 303 WIHNTHGSG-IGKPIEFATHTSSRNARNVSOAWRHALPFTKILRVQVCAGALQLLGYRP 361
DB 408 falmtsgssskp---fvvsarnatqaanawrtaltfqgikvvefcyqpmavlgvyr 464
QY 362 VYSADQQRDLTLDLV 376
DB 465 vnspeevkdsktll 479

RESULT 12


```

AAW52863
ID AAW52863 standard; Protein; 479 AA.
XX
XX AAW52863;
XX
XX 30-JUN-1998 (first entry)
XX
XX Glycosaminoglycan sulphotransferase protein.
DE
XX
XX Glycosaminoglycan sulphotransferase; C6ST; N-acetylgalactosamine;
KW galactose; glycosaminoglycan; chondroitin; keratan sulphate; diagnostic.
KW
XX Homo sapiens.
OS
XX
XX EP821066-A1.
PN
XX
XX 28-JAN-1998.
PD
XX
XX 22-JUL-1997; 97EP-0305476.
PF
XX
XX 24-JUL-1996; 96JP-0195063.
PR
XX
XX (SEKG ) SEIKAGAKU KOGYO CO LTD.
PA
XX
XX Fukuta M, Habuchi O;
PI
XX
XX WPI; 1998-102622/10.
DR
XX
XX Human glycosaminoglycan sulpho:transferase protein - catalyses
PT 6-sulphation of chondroitin to produce chondroitin sulphate for use
PT in pharmaceutical(s)
PT
XX
XX Claim 3; Page 21-23; 27pp; English.
XX
XX The human glycosaminoglycan sulphotransferase (C6ST) is a type II
CC membrane protein derived from foetal brain tissue with a molecular weight
CC of 50-55 kD. The protein transfers sulphate groups from a donor to the
CC N-acetylgalactosamine residue or galactose residue of glycosaminoglycan.
CC C6ST displays substrate specificity transferring the sulphate group to
CC the hydroxyl group position at C-6 of the N-acetylgalactosamine residue
CC of chondroitin and the hydroxyl group position at C-6 of the galactose
CC residue of keratan sulphate. The protein is useful for studying the
CC function of chondroitin sulphate and can provide chondroitin sulphate
CC that may be useful in pharmaceuticals. Both the protein and DNA might
CC be useful for treating or diagnosing diseases attributable to low C-6
CC sulphation of chondroitin N-acetylgalactosamine residues.
XX
XX Sequence 479 AA;
XX
XX Query Match 28.2%; Score 577.5; DB 19; Length 479;
XX Best Local Similarity 35.7%; Pred. No. 2.6e-54;
XX Matches 132; Conservative 65; Mismatches 134; Indels 39; Gaps
XX
XX 32 PSSPAGGEDRVHVLVLSWSRSGSFLGOLFQHPDVFLMEPAHHVWTLTSL----QGSAA 87
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 123 pprpavagprthvlmatrrcgssvfgeffnggniflyfeplwhiertvtsfegpganaa 182
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 88 TLHMAVRDLMSRIFLCMDVFPDAY---MPQSRNLSAFENWATSRALCSPACSAFPRGTI 144
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 183 gsalvyrdvklqfclcdlylvlehitplpedhltqfmgrrssrlcedpvcctpfvkvkf 242
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 145 SKQDVCKTLCTRFQFSIAREACRSYSHVVLKEVRFNQLVLYPLSDPALMLRIHVLRD 204
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 243 ekychknrrcgplnvtiaaeacrrkemmaikavirgleflqlpLaedprldrlvqlvrd 302
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 205 PRAVLRSEAAGPILARDNGIVLQ---TNGKWE-----ADPHLRLIRVCKSHVRIA 254
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 303 pravlasrmtva-----fagkyktwkwlddegdglreeevqrlrgnces-irls 351
XX
XX 255 EAAVTKPPFLRGYRLVRVEDIAREPLAEIRALYATFTGLTLPQLEAWHTNTH----G 310
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 352 aelglrpalwrlrgmylrvyedvargplqkaremypfaglpftrpqvedw.lqktgaahdg 411
XX

```

DNA encoding chondroitin 6-sulphotransferase ~ for recombinant production of C6ST, for use in industrial processes

Claim 2; Page 18-20; 30pp; English.

Chick embryo chondrocyte chondroitin 6-sulphotransferase (C6ST) (AAW06480) is capable of catalysing the transfer of a sulphate group from 3'-phosphoadenosine 5' phosphosulphate to the hydroxyl group at the C-6 position of the N-acetylglactosamine residue of chondroitin. Its amino acid sequence was deduced from a cDNA clone (AAU45037) isolated from a chick embryo chondrocyte cDNA library. Isolation of the cDNA allows the mass prodn. of C6ST in transfected host cells. Recombinant C6ST polypeptides are useful for analysing the activities of chondroitin sulphate and for modifying its function, and for raising antibodies.

Sequence 458 AA;

Query Match 26.4%; Score 540.5; DB 18; Length 458;
Best Local Similarity 34.3%; Pred. No. 2.8e-50;
Matches 124; Conservative 66; Mismatches 141; Indels 31; Gaps

QY 36 AGGEDRVHVLVLSWSRSGSFLGQLFSQHPDVFYLMPEAWHV--WTTLSQGSAAATLHMAY 93
DB 107 aaapeprhvlmatrtgssfgfngqgnifyleplwhiertvttfepganavgsal 166
QY 94 --RDLMRSIFLCMDVFADYM---PSQRNLSAFFNWATRSLCPPACSAFPRGLISKQD 148
DB 167 vyrdvlgllldlyallesfisapaeehitaalfirgsshslceepvcvtpslkkvfekyh 226
QY 149 VKCTLCTRPQFSLARACRSYSHWLVKEYRFENLQVLYPLLSDPALNLRIVHLVRDPRAV 208
DB 227 cknrcpelnitlaaacrrkqdmakltvirqlqleflqlpaedprldrlriqlvrdrav 286
QY 209 LRSREAGPILARDNGVIGLTNGKWVE-----ADPHLLIREVCNSHVRIAEATL 259
DB 287 lvsrmva-----fskgyeswkwkaageaplqedevgrlrgnces-irlsaelgi 335
QY 260 KPPFLRGRLRYRFEDLAREPLAETRALYAFTGLTLTQLEAWTHNIITHGSGICKPIEA 319
DB 336 iqpwlrgrymlvryedvaraprkalemyrfagibtpqvvewirantcqp---qdsng 392
QY 320 FHTSSRNARNVSQAWRHALPFTKILIRVQEVCA GALOLLGYRPVYSADQQORDITLDLVLPR 379
DB 393 iystgkossefekewrifsipfkiaqvvgdacepamlfyglylassaqeltnrsls!-lee 451
QY 380 GP 381
DB 452 gp 453

RESULT 14
AAW61100
ID AAW61100 standard; Protein; 411 AA.
XX AC AAW61100;
XX DT 14-SEP-1998 (first entry)
XX DE Keratan sulphate 6-sulphotransferase.
XX KS Gal6ST; keratan sulphate 6-sulphotransferase;
KW chick chondroitin 6-sulphotransferase; C6ST; phage Lambda; hybridization;
KW expression vector; COS-7 cells; Bluescript plasmid; galactose;
KW keratan sulphate; chondroitin.
XX CS Homo sapiens.
XX FN EP845533-A2.
XX Q3-JUN-1998.

Search completed: September 20, 2002, 19:58:07
Job time: 161 sec

		91674: gap of unknown length	91675	91676	91677	91678	91679	91680	91681	91682	91683	91684	91685	91686	91687	91688	91689	91690	91691	91692	91693	91694	91695	91696	91697	91698	91699	91700	91701	91702	91703	91704	91705	91706	91707	91708	91709	91710	91711	91712	91713	91714	91715	91716	91717	91718	91719	91720	91721	91722	91723	91724	91725	91726	91727	91728	91729	91730	91731	91732	91733	91734	91735	91736	91737	91738	91739	91740	91741	91742	91743	91744	91745	91746	91747	91748	91749	91750	91751	91752	91753	91754	91755	91756	91757	91758	91759	91760	91761	91762	91763	91764	91765	91766	91767	91768	91769	91770	91771	91772	91773	91774	91775	91776	91777	91778	91779	91780	91781	91782	91783	91784	91785	91786	91787	91788	91789	91790	91791	91792	91793	91794	91795	91796	91797	91798	91799	91800	91801	91802	91803	91804	91805	91806	91807	91808	91809	91810	91811	91812	91813	91814	91815	91816	91817	91818	91819	91820	91821	91822	91823	91824	91825	91826	91827	91828	91829	91830	91831	91832	91833	91834	91835	91836	91837	91838	91839	91840	91841	91842	91843	91844	91845	91846	91847	91848	91849	91850	91851	91852	91853	91854	91855	91856	91857	91858	91859	91860	91861	91862	91863	91864	91865	91866	91867	91868	91869	91870	91871	91872	91873	91874	91875	91876	91877	91878	91879	91880	91881	91882	91883	91884	91885	91886	91887	91888	91889	91890	91891	91892	91893	91894	91895	91896	91897	91898	91899	91900	91901	91902	91903	91904	91905	91906	91907	91908	91909	91910	91911	91912	91913	91914	91915	91916	91917	91918	91919	91920	91921	91922	91923	91924	91925	91926	91927	91928	91929	91930	91931	91932	91933	91934	91935	91936	91937	91938	91939	91940	91941	91942	91943	91944	91945	91946	91947	91948	91949	91950	91951	91952	91953	91954	91955	91956	91957	91958	91959	91960	91961	91962	91963	91964	91965	91966	91967	91968	91969	91970	91971	91972	91973	91974	91975	91976	91977	91978	91979	91980	91981	91982	91983	91984	91985	91986	91987	91988	91989	91990	91991	91992	91993	91994	91995	91996	91997	91998	91999	92000
91674:	gap of unknown length	91675	91676	91677	91678	91679	91680	91681	91682	91683	91684	91685	91686	91687	91688	91689	91690	91691	91692	91693	91694	91695	91696	91697	91698	91699	91700	91701	91702	91703	91704	91705	91706	91707	91708	91709	91710	91711	91712	91713	91714	91715	91716	91717	91718	91719	91720	91721	91722	91723	91724	91725	91726	91727	91728	91729	91730	91731	91732	91733	91734	91735	91736	91737	91738	91739	91740	91741	91742	91743	91744	91745	91746	91747	91748	91749	91750	91751																																																																																																																																																																																																																																																										

```

Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coville, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy-Garroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Gaster, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lien, C., Liu, J., Liu, W.,
Loulis, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapa, P., Martin, R., McLeod, M.P., Meador, M.,
Martinez, E., Massey, E., Mawhiney, E., Miner, Z., Mitchell, T., Mohabbat, K.,
Mei, G., Metzker, M., Miner, G., Miner, G., Miner, Z., Newton, J., Newton, N.,
Morgan, M., Morris, S., Moser, M., Nguyen, N., Nickerson, E., Nwokwenkwo, S.,
Nguyen, A., Nguyen, N., Nguyen, N., Oquendo, R., Pace, A., Payton, B.,
Ogih, M., Okwuonu, G., Oquendo, R., Oquendo, R., Oquendo, R., Oquendo, R.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shooshtari, N.,
Sisson, I., Sodergren, E., Sonaik, P., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmami, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczky, R., Woodan, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 139252)
Worley, K.C.
Direct Submission
Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:15627284.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GCW
Center clone name: CH230-8L17
Center clone name: Summary Statistics
----- Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 96308 bases at least Q40
Consensus quality: 106296 bases at least Q30
Consensus quality: 113927 bases at least Q20
Estimated insert size: 91054; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-gel estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 6038: contig of 6038 bp in length
* 6039 6138: gap of unknown length

```

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 1740)
TITLE Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue
JOURNAL Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)
MEDLINE 99423499
REFERENCE 2 (bases 1 to 1740)
AUTHORS Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
TITLE Direct Submission
JOURNAL Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
source Location/Qualifiers
1..1740
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/map="8E1"
/tissue_type="intestine"
/note="Possibly involved in biosynthesis of mucosal ligands for L-selectin"
/product="N-acetylglucosamine 6-O-sulfotransferase"
/gene="I-GlcNAc-6-ST"
/gene="I-GlcNAc-6-ST"
/note="Possibly involved in biosynthesis of mucosal ligands for L-selectin"
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD56003.1"
/db_xref="GI:591712"
/translation="MRLPRSSFTVMSLLVQVGIILVLSRVSPSAGLGERVHVL VLSWRGSSVQGLFSDHPDYFYLMEPAWVDTLQGSAPALHAWRLDLRSVFLC DMVDYSLVPMRNLSDLFQWAVSRALCPSPVCEAFARGNISSEVKPLCATRPFL ACESSYSHVLEKVRFFNLVPLSDPALNRIHLVDPRVLRSDRTAKAL ARDGLVIGTNGWEADPRLVNVEYRSHVRIAEALHKKPPFDQDRYLRVREDL ARDELTVIRLYEATPGLTLPQLTWLNHTHSGSPGARREAPKPTSRDALSVSOAWR HTLPFAKTRRYOELCGGALQLGYSRVHSELEQRLSLDLILPRGMDSPKWSSTEQ PES"
BASE COUNT 365 a 513 c 491 g 371 t
ORIGIN

Query Match 63.5%; Score 745.2; DB 10; Length 1740;
Best Local Similarity 78.0%; Pred. No. 2.2e-90;
Matches 910; Conservative 0; Mismatches 253; Indels 3; Gaps 1;

QY 1 agtgagctccacgcttctccagacagacagtcagtcagtcctctctgagcagacac 60
DB 405 ATGGGGTACCCGGTTTCCAGCACTGTCTTACCTGATGGAGCGGGCTTGGCAGCTGGGATACG--ACT 461
QY 61 tgcctctgctcttcatcatctccggccagggccctcatcccccagcggcgagagat 120
DB 462 GGCATCTCTGTTCTTCCTCGGCAAGTGCATGTCCTCCAGCAGGCGCTTGGGGAG 521
QY 121 cgtatgacagctgctgctgctcctggtggtggtggtggtggtggtggtggtggt 180
DB 522 CGHGTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
QY 181 ttcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 240
DB 582 TTCAGCAACACCCCGATGTCTTACCTGATGGAGCGGGCTTGGCAGCTTGGGATACG 641
QY 241 cgtgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 300
DB 642 TTGTCAGGAGGAGGCGCCCGCACTCCACATGCGCGGCGGTGACCTGATCCGCTCAGTG 701

QY 301 ttttgtgacatggacgtgtttgatccctacatgccacagagccgcaaacctgtccgc 360
DB 702 TTCCTATGGCATGGAGGTATTTGATGCTTACCTGCTGCGCGCCGCAACATCTGGAT 761
QY 361 ttttcaactgggcaaacagcgcgcgtgtgtctgcccgcgcgcgcgcgcgcgcgcgc 420
DB 762 CTCCTTCCAGTGGCGGTGAGCGCGCATFTGTCTACCTCCGGTCTCGAGGCTTCGCT 821
QY 421 cagggcacatcagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
DB 822 CGTGGCAACATCAGCAGCAGGAGGTGTGTAAGCTCTGTGGCAACCGCGCTTCGCG 881
QY 481 ctggccggaggc 540
DB 882 CTGGCTCAGGAAGCTGCAGCTCTATAGTACCTGTCTGTCACAGGAGGTGCTCTT 941
QY 541 aactgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 600
DB 942 AACCTACAGGTCTCTACCCGCTGCTCAGCAGCACTGCTGCTCAACCTGCTGCTG 1001
QY 601 ctggcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 660
DB 1002 CTAGTGGCGAGCCCGCGCGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1061
QY 661 cgcgcaacgcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 720
DB 1062 CGGGACAATGGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1121
QY 721 cgcctgattcgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 780
DB 1122 CCGGTGGTCAAGCAGGATGATGCGGCGCATGTGCGCATGCGCAGAGGAGCGCTTGC 1181
QY 781 cgcgcacccctctcgc 840
DB 1182 CGCGCGCTCTTTCGAAGATCGCTACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1241
QY 841 cgcgtggcagagatccgc 900
DB 1242 CCACCTACCGTAATCCGTTGAACTCTATGCTTACCGCGCTGCTGCTGCTGCTGCTGCT 1301
QY 901 gaggcctggatcccaacatcacccacccgggtcggggagtcggcaagccaatcgaggc 960
DB 1302 CAGACTTGGATCCCAATATCACATGCTTACAGGGCAGCGCGCGCTGGAAGCTTTC 1361
QY 961 catctctgttagaagtcgc 1020
DB 1362 AAGCCACATCAGGAGTGGCTCAGTGTATCCAGGCGCTGCGCGCTGCTGCTGCTGCT 1421
QY 1021 actaagatcctgc 1080
DB 1422 GCCAAGATTCGCGGGTCCAGGAACCTGTGCGGGGTGCACTGCACTGCTGCTGCTGCT 1481
QY 1081 cctgtactctcggaccagcagcagcagcagcagcagcagcagcagcagcagcagcag 1140
DB 1482 TCTGTGCAATCGGAGCTTGAGCAAGGACCTCTCTGTGACCTCTCTGCTGCTGCTGCT 1541
QY 1141 ccagaccattcagctgggcatcgcc 1166
DB 1542 ATGGACAGTTTCAAGTGGGATCGTC 1567

RESULT 14
AF176840
LOCUS
DEFINITION Mus musculus intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-ST) mRNA, complete cds.
ACCESSION AF176840
VERSION AF176840.1
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1989 bp mRNA linear ROD 22-SEP-1999

[illegible][illegible]

Fri Sep 27 20:49:20 2002

QY	1141	ccagaccacttcagctgggcatc 1163	
Db	1830	CTGACGGCTTCACCTGGGCATC 1852	
RESULT 10			
AF280086		3786 bp	linear
LOCUS			PRI 20-FEB-2001
DEFINITION		Homo sapiens N-acetylglucosamine 6-O-sulfotransferase GST-4beta	
		mRNA, complete cds.	
ACCESSION		AF280086	
VERSION		AF280086.1	GI:12060803
KEYWORDS		human.	
SOURCE		Homo sapiens	
ORGANISM		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 3786)	
AUTHORS		Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and Rosen,S.D.	
TITLE		Chromosomal localization and genomic organization for the galactose/ N-acetylglactosamine/N-acetylglucosamine 6-O-sulfotransferase gene family	
JOURNAL		Glycobiology 11 (1), 75-87 (2001)	
MEDLINE		21096027	
REFERENCE		2 (bases 1 to 3786)	
AUTHORS		Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and Rosen,S.D.	
TITLE		Direct Submission	
JOURNAL		Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA.	
FEATURES		Location/Qualifiers	
		1..3786	
		/organism="Homo sapiens"	
		/db_xref="taxon:9606"	
		/chromosome="16"	
		/map="16q23.1"	
		1..3786	
		/gene="GST4beta"	
		/note="clusters with the intestinal N-acetylglucosamine 6-O-sulfotransferase (GST4alpha) gene in GenBank Accession Numbers AF176838 and AF176839"	
		189..1376	
		/gene="GST4beta"	
		/note="similar to N-acetylglucosamine 6-O-sulfotransferase GST-4alpha"	
		/codon_start=1	
		/product="N-acetylglucosamine 6-O-sulfotransferase	
		GST-4beta"	
		/protein_id="AAG48244.1"	
		/db_xref="GI:12060804"	
		/translation="MWLPVRSSTAVTALLAQTFLLFLVSRPGSPSPAGGEARVHVL VLSWRSGSFFVQLFNQHPDVFYLMFPAMHVTTLSCGSAATLBMAYRDLVRSVELC DMQVDFAYLPWRNLSDLQFQVSRALCSPACAFPRGAISSSAVCAFLCARQSFLL AREACRSYSHVLEKREFNLOVLYPLSDPALNRIYHLDVDPVAVLRREQAKAL ARDNGVLVGTNGTWEADPGLRVVRVCRSHVRIAEATLKPFPFLRGRLVRFEDL AREPLAEIRALYAFGLSLTQLEAHINHTHSGSPGARREAFKTSNNALNVSQAWR HALPFAKIRRYQELCAGALQLLGYRVPYSEDEQRNALDLVLPGLNGFTWASSTASH PRN"	
		3106..3111	
		/gene="GST4beta"	
		/note="alternative signal"	
		3734..3739	
		/gene="GST4beta"	
		848 a 1030 c 1104 g 804 t	
		polyA_signal	
		polyA_signal	
		BASE COUNT 848 a 1030 c 1104 g 804 t	
		ORIGIN	
		Query Match 78.8%; Score 924.6; DB 9; Length 3786;	
		Best Local Similarity 87.8%; Pred. No. 2.6e-114;	
		Matches 1021; Conservative 0; Mismatches 139; Indels 3; Gaps 1;	
		QY 1 atgtggtgcgcaggttctctccagcaagcagtgacagtgcctctctctctgggcacagaccacc 60	

[illegible]

Qy	1081	cctgttactctgcggaccacgcagctgtgacctcaacctgcctgactgctgctgcacagaggc	1140
Db	1078	CCTGTGACTCTTGAGGACGAGCAGCGCAACCTGCCCTTGATCTGGTGTGCTGCCAGGCG	1137
Qy	1141	ccagaccacttcagctgggcac	1163
Db	1138	CTGAACGCTTCACTTGGGCATC	1160
RESULT	9		
LOCUS	AF219990	2544 bp	linear
DEFINITION	Homo sapiens corneal N-acetylglucosamine-6-O-sulfotransferase		
ACCESSION	AF219990		
VERSION	AF219990.1	GI:11023145	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 2544) Akama,T.O., Nishida,K., Nakayama,J., Watanabe,H., Fujiwara,T., Ozaki,K., Kinoshita,S., Shimomura,Y., Tanigami,A. and Fukuda,M.N.		
TITLE	Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene		
JOURNAL	Nat. Genet. 26 (2), 237-241 (2000)		
MEDLINE	20472330		
PUBMED	11017086		
REFERENCE	2 (bases 1 to 2544) Akama,T.O. and Fukuda,M.N.		
AUTHORS	Direct Submission		
TITLE	Submitted (29-DEC-1999) Glycobiology Program, The Burnham		
JOURNAL	Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES	Location/Qualifiers		
source	1..2544		
gene	/organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="16" /map="16q22" 1..2544 /gene="CHST6"		
CDS	/note="responsible for macular corneal dystrophy" 693..1880 /gene="CHST6" /function="transfers sulfate to 6-OH of GlcNAc in keratan sulfate" /note="C-GlcNAc6S"		
	/codon_start=1 /product="corneal N-acetylglucosamine-6-O-sulfotransferase" /protein_id="AAG26325.1" /db_xref="GI:11023146" /translation="MWLPRVSTAVTALLAQTFLLFLVSRPQSPSSPAGEARVHVLVDNSRSGSFFVQGLFNGRPDVFYLMFAFWHTLISQGSAAITLHMAVROLIVRSVFLCDMDVDAILPWRNLSDLQWASRALCSPSPACAFPRGATISSEAVKPLCAQGSFTLAREACRSVVVLEKEVFNFIQVYPLLSDPALNRLVHIVRPVLRSGEQTAKALARDNGIVLGTNGVWEADPGLRVVRECVRSHTVTAEAATLKPPFLGRVLYRPFDELAREPLAEIRALYATFGLSLTQPLBAILNHTHCGSGCARREAPKTSRNALNVSQAWHALPFAKIRRYQELCAGNALQLLGYRPFYSEDEQRNALDLVLRGLNGFTWASSTASHPRN"		
BASE COUNT	460 a 799 c 733 g 552 t		
ORIGIN			

```

Query Match      78.8%   SCORE 924.6; DB 9; Length 2544;
Best Local Similarity 87.8%; Pred. No. 2.8e-114;
Matches 1021; Conservative 0; Mismatches 139; Indels 3; Gaps 1;

Qy 1 atgttggtgcacagcgtttctccagcacagtcagcagtcctctcctgagcacacacc 60
    |||||
Db 693 ATGTGGTGGCGGCGCTCCAGCACAGTCAGCGCGCTCCCTGTGGCGCAG-ACC 749
    |||||

```

[illegible]

Qy	901	gagcctggatccacaacatcaccacccggtcgggggtacgcaagccaatcgaagccttc	980
Db	18848	GAGGCTGGATCCACAATACCCACGGGTGGGGATCGGCAAGCCAACTGAGGCGCTTC	188507
Qy	951	catacttcgttaggaatgcgcgaactccacgctccggccagcgcgttgcccttc	1020
Db	188508	CATACTTCGTAGGAATGCGCGCACTGCCAGGCTGGGCCACGCGTTGCCCTTC	188567

[illegible]

QY	1141	ccagaccacttcagctgggcatcgctgcaciga	1173
DB	188688	CCAGACCACTTCAGTGGGCATCGCCTGACTGA	189720
RESULT	8		
AX327330		1647 bp	DNA
LOCUS		AX327330	RefSeq NC_017946.8
			linear
			PAT 07-JAN-20

DEFINITION	Sequence 15 from Patent WO01/5760.
ACCESSION	AX327330
VERSION	AX327330.1 GI:18097876
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS	Policky, J.L., Hatala, A., Burford, N., King, A., Patterson, C., Das, Tribouley, C., Yao, M.G., Yue, H., Tang, Y.F., Patterson, C., Das, Sanjanwala, M.S., Gandhi, A.R., Reddy, R., Khan, F.A., Baughn, M.R., Ramkumar, J., Griffin, J.A. and Au-Young, J.
TITLE	Drug metabolizing enzymes
JOURNAL	Patent: WO 0179468-A 15 25-OCT-2001; Incyte Genomics, inc. (US)
FEATURES	Location/Qualifiers 1 .1547

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/note="Incyte ID No: 747277CB1"
287 a 560 c 501 g 299 t
BASE COUNT
ORIGIN
78.8% Score 924.6; DB 6; Length 1647;
Query Match
87.8% Pred. No. 3.2e-114;
Posterior Similarity

```

Best Local Alignment: 1021; Conservative: 0; Mismatches: 139; Indels: 3; Gaps: 0

* as soon as it is available and the accession number will
* be preserved.

```

1 1084: contig of 1084 bp in length
* 1184: gap of unknown length
* 1185: contig of 1211 bp in length
* 2395: gap of unknown length
* 2495: gap of unknown length
* 2496: contig of 1785 bp in length
* 4281: gap of unknown length
* 4380: gap of unknown length
* 4381: contig of 1778 bp in length
* 6159: gap of unknown length
* 6259: contig of 2354 bp in length
* 8613: gap of unknown length
* 8713: gap of unknown length
* 11625: contig of 2913 bp in length
* 11725: gap of unknown length
* 11726: contig of 1531 bp in length
* 13257: gap of unknown length
* 13357: contig of 3123 bp in length
* 16480: gap of unknown length
* 16579: gap of unknown length
* 19343: contig of 2764 bp in length
* 19443: gap of unknown length
* 23596: contig of 4153 bp in length
* 23597: gap of unknown length
* 23697: contig of 5485 bp in length
* 29181: gap of unknown length
* 29281: gap of unknown length
* 35079: contig of 5798 bp in length
* 35179: gap of unknown length
* 39189: contig of 4010 bp in length
* 39289: gap of unknown length
* 48543: contig of 9254 bp in length
* 48544: gap of unknown length
* 90448: contig of 41805 bp in length
* 90449: gap of unknown length
* 90549: contig of 68524 bp in length.

FEATURES
    source
        1..159072
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CTD-2113H21"
            /clone_lib="CalTech human BAC library D"
BASE COUNT 43207 a 36742 c 36517 g 41088 t 1518 others
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 1173; DB 2; Length 159072;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgtgtgctgccaagttctccagcaagacagtgacagtgctctctctctgacacagaccacc 60
Db 61400 ATGTGCTGCCACGGTTCTCCAGCAAGACAGTGACAGTGCTCTCTCTGACACAGACCACC 61459

Qy 61 tgcctctgtcttcttcacatctcccgccagggccctccatccacagccggcgagagat 120
Db 61460 TGCCTCTCTCTTTCATCATCTCCGGCCAGGGCCCTCATCCCCAGCCGGCGGAGAGAT 61519

Qy 121 cgtgtgacgtgctggtgtctctctgtggtgctggtggtcctctcttcttgggcagctc 180
Db 61520 CGTGTGACGTGCTGGTGTCTCTCTCTGCGCTCGGGCTCATCTCTTGGGCCAGCTC 61579

Qy 181 ttacagcagcaccacagctcttctactgatgatgagcccgctggtggtgacacc 240
Db 61580 TTACGCCAGCACCCCGAGCTTTCTACCTGATGAGCCCGCTCATCCCCAGCCGGCGGAGAT 61639

Qy 241 ctgtcgaggagcagcgccagcgtgcacatgcccgtgcccgcagcctgatcgctctatc 300
Db 61640 CTGTGCGAGGSCACGGCGGCAACGCTGCACATGCGCTGCGGACCTGATGCGCTCTATC 61699

Qy 301 tttttgtgacatgagcgtgtttgatgctctacatgcccagagccgaacacctccgcc 360
Db 61700 TTTTGTGACATGGACGCTTTGATGCTCTACATGCCACAGACCGAACCCTGTCGCC 61759

Qy 361 tttttcaactgggcaacagccgcgctgtgtcgcgcgcgcgcgcgcgcgcgcgcgcgc 420

```

```

Db 61760 TTTTTCAACTGGGCAACGAGCGCGCTGTGCTGCGCCCGCCCTGCGAGCGCTTCC 61819
Qy 421 cgaagcaccatcagcaagcaggaacatgcaagacactgtgcaagcagcagcattcagc 480
Db 61820 CGAGCACCATCAGCAAGCAGGACGATATGCAAGACACTGTGCAACGCGCAGGCATCAGC 61879
Qy 481 ctggcccgaggagcgtccgtctccctacagccacgtggtgctcaagagagtgctctctc 540
Db 61880 CTGGCCCGGAGGCGCTGCGCGTCTTACAGACCCACGTTGTCTCAAGAGAGTGCGTCTTC 61939
Qy 541 aacctgaagtgctctaccgcgtctcagcgaccccccgcgtccacacctgcgcatgtgcac 600
Db 61940 AACCTGACGTTGCTCTACCGCTGCTCAGGACACCCCGCGCTCAACCTGCGCATGTGAC 61999
Qy 601 ctggtgcgcacccgcggccgtgctcggtctcccgaggagcggtggcccgatactggca 660
Db 62000 CTGGTGGCGAGCCCGCGGCGGTGCTGCGCTCCGCGGAGGCGGCGCGGCGGCGGAT 62059
Qy 661 cggacacacgcatgctgtggccacacggaagtggtgggtggagcgccacctcacctg 720
Db 62060 CGCGACAACGCGATCGTGTGGGCACCAACGCGCAAGTGGGTGGAGCGCGGCGGCGGAG 62119
Qy 721 cgcctgattcgcaggtgtgcgcagcgaogtgcgcacgcgcgcgcgcgcgcgcgcgcgc 780
Db 62120 CGCCTGATTCGAGGTTGTGCGCAGCACCGTGGCATCGCCGAGGCGCGCACACTCAAG 62179
Qy 781 ccgccacctctctgcgcgcgcacacacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 840
Db 62180 CCGCCACCTTCTGCGCGCGCGCTACCGCTGTGTGGCTTCGAGGACCTTGGCGCGGAG 62239
Qy 841 ccgctggcagagatccgcgcacctctacgccttcacgcgcgcgcgcgcgcgcgcgcgcgc 900
Db 62240 CCGCTGCGCAGATCCGCGCACTTACGCGCTTACGCGGCTTGACCCCTACGCCACAGCT 62299
Qy 901 gaggcctggatccacacatcccccgggtgggtgggtgggtgggtgggtgggtgggtgggt 960
Db 62300 GAGGCTGGATCCACACATCACCCACGGGTGGGGATCGGCAAGCAATCGAGGCGCTTC 62359
Qy 961 catacttcctaggaatcgcgcaacgtctccagcgcgcgcgcgcgcgcgcgcgcgcgcgc 1020
Db 62360 CATACTTCCTTAGGAATCGCGCAACGCTCCAGGCTGGCGGCTGGCGGCGGCGGCTTC 62419
Qy 1021 ataagatcctgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1080
Db 62420 ACTAAGATCCTGGCTGCGAGGAGGTGTGCGCGCGCGCGCTGCAGCTGTGGGCTACCG 62479
Qy 1081 cctgtgactctggacacagcagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
Db 62480 CCTGTGACTCTGGGACACAGCAGGTGACCTCACCTGGATCTGTGTGCTGCCACGAGGC 62539
Qy 1141 ccagaccacttcagctgggcacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1173
Db 62540 CCAGACCACTTACGTTGGCATCGCTGACTGA 62572

RESULT 7
AC025287 194832 bp DNA linear PRI 26-JAN-2002
LOCUS Homo sapiens chromosome 16 clone RP11-490B18, complete sequence.
DEFINITION AC025287
ACCESSION AC025287
VERSION AC025287.8 GI:18376863
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 194832)
AUTHORS DOE Joint Genome Institute.

```

Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES
Source
1. .157337
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-77K12"
BASE COUNT 45295 a 36632 c 36308 g 39102 t
ORIGIN

Query Match 100.0%; Score 1173; DB 9; Length 157337;
Best Local Similarity 100.0%; Pred. No. 9,3e-146;
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgtgctgcacgggtctccagcaagacagtgacagtgtctctctctggcacagaccacc 60
Db 144585 ATGTGGCTGCCAGGTTCTTCAGCAACAGTGCAGTGTCTCTCTGGCACAGACCACC 144526
QY 61 tgcctcctgtcttctcatcattctccggccagggccctccatcccccagccgagggagat 120
Db 144525 TGCTCTGCTCTTCATCATCTCCGCGCAGGGCCCTCATCCCGAGCGCGCGAGGAT 144466
QY 121 cgtgtcaagctgtggtgtctctctctgtggtggtcctctctctctctctctctctct 180
Db 144465 CGTGTGCAGCTGCTGGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 144406
QY 181 ttcagcagaccccgagctcttctacctgatgagccggtggtggtggtggtggtggtg 240
Db 144405 TTCAGCCAGACCCCGAGCTCTTCTACTGTAGTGCAGCCGCGTGCATGTGGACACC 144346
QY 241 ctgtcgcagggcagc 300
Db 144345 CTGTGCAGGCGAGCGCGGCAACGCTGCACATGGCGGTGCGCGACCTGATGCGCTATC 144286
QY 301 tttttgtgcacatggacgtgtttgatcctatcatgcacagagcgaacgtgccgcc 360
Db 144285 TTTTGTGCGACATGGACGTGTTTGTATGCTACATGCCAGAGCGGAACCTGTCCGCC 144226
QY 361 tttttcaactggcagcagc 420
Db 144225 TTTTCAACTGGGCAACGAGCGCGCGCTGTCTGTCCGCGCGCTGCAGCGCTTCC 144166
QY 421 cagggcaccatcagcaagcagcagcagcagcagcagcagcagcagcagcagcagcagc 480
Db 144165 CGAGGCACCATCAGCAAGCAGGACGCTATGCAAGACACTGTGCAGCGCGGCGAGCTCAGC 144106
QY 481 ctggcccgaggagcgtgcgcgtcctacagcagcgtggtgtctcaaggagtgcgctcttc 540
Db 144105 CTGGCCCGGAGGCGCTGCGCGCTCTACAGCCACGCTGCTCAAGGAGGTGGCTTCTTC 144046
QY 541 aacctgcaggtgtctacccgtgtcctcagcagcccgctcctcaacctgcgcgtcgtgac 600
Db 144045 AACCTGCAGGTGCTACCCGCTGTCTCAGGACCCCGCGCTCAACCTGCGCATCGTGAC 143986
QY 601 ctggtgcgcgaccccgccggtgtgctgctcctccggagcgggcgccgacatactggca 660
Db 143985 CTGGTGCAGCGACCCGCGGCGCGTGTCTGCTCTCCGGGAGGCGCGCGCGGCTACTGGCA 143926
QY 661 cgcgcaacagcgatcgtgtgtggcacaacagcgcaagtgtggtgagccgacccctcactg 720
Db 143925 CCGCAACAGCGCATCTGTGTGGGACCAACAGGCAAGTGGGTGGAGCCGCGGACCTTCACCTG 143866
QY 721 cgcctatcgcaggtgtgtccgcagccagcagcagcagcagcagcagcagcagcagcagc 780
Db 143865 CGCCTGATTCGCGAGGTGTGCGCGAGCCAGCGTGGCATCCCGAGCGCGCGACACTCAAG 143806
QY 781 ccgccaccctctctgc 840

Db 143805 CCGCCACCCCTTCTGCGCGCGCGCTACCGCTGTGGTGGCTTCGAGGACCTGGCGCGGAG 143746
QY 841 ccgctggcaagatccgc 900
Db 143745 CCGCTGGCAGAGATCCGCGCACTTACGCTTTCACCGGCGCTGACCCCTACGCCACAGCTC 143686
QY 901 gaggcctggatccacaacatcacccacggtcggggagtcgcaagcgaatcgaggccttc 960
Db 143685 GAGGCTGGATCCACAACATCACCCAGGGTGGGGATCGCAAGCAATCGAGGCGCTTC 143626
QY 961 catacttcgtctaggaatgc 1020
Db 143625 CATACTTCGTCTAGTAATGCGCAACGCTCTCCAGGCGCTGGCGCGCTGGCCCTTC 143566
QY 1021 actaagatcctgc 1080
Db 143565 ACTAAGATTCGTCTGCGCTGCGAGAGGTGTGCGCGCGCGCTGCGAGCTGTGGTGTGCGCAGGCG 143506
QY 1081 cctgttactctgc 1140
Db 143505 CCTGTGTACTCTGCGGACCGAGCGTGCACCTCACCTTGTGATCTGGTGTGCGCAGGCG 143446
QY 1141 ccagaccacttcagctggc 1173
Db 143445 CCAGACCATTTCAGTGGGCATCGCTGACTGA 143413

RESULT 6
AC026419 159072 bp DNA linear HTG 20-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2113H21, WORKING DRAFT
DEFINITION SEQUENCE, 16 unordered pieces.
AC026419 GI:3699674
VERSION AC026419.3
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159072)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 159072)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (22-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Apr 20, 2001 this sequence version replaced gi:7711893.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 671130, H460
Center clone name: CITB-H1_2113H21

Summary Statistics
Consensus quality: 140723 bases at least Q40
Consensus quality: 148960 bases at least Q30
Consensus quality: 151290 bases at least Q20
Estimated insert size: 98000; pulse field gel estimation
Estimated insert size: 157572; sum-of-contigs estimation
Quality coverage: 7.39 in Q20 bases; pulse field gel estimation
Quality coverage: 4.59 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence

repeat_region	23344..23400
	/rpt_family="MaLR"
repeat_region	/rpt_type=dispersed 23433..23456
	/rpt_type=tandem
repeat_region	/rpt_family="Simple_repeat" complement(23457..23720)
	/rpt_type="Alu"
repeat_region	/rpt_type=dispersed complement(23789..24140)
	/rpt_family="MER21-group"
repeat_region	/rpt_type=dispersed 24141..24287
	/rpt_family="MaLR"
repeat_region	/rpt_type=dispersed complement(24308..24420)
	/rpt_family="MaLR"
repeat_region	/rpt_type=dispersed complement(24467..24528)
	/rpt_family="MER1-type"
Query Match 100.08; Score 1173; DB 9; Length 71503;	
Best Local Similarity 100.08; Pred. No. 1.e-147;	
Matches 1173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1	atgtggctgccagcgttctccagcaagacagtgacagtcgctoctoctctgcacagaccacc 60
DG 19030	ATGTGGCTGCCAGGTTCTCCAGCAAGACAGTGACAGTGCCTCCTCGCCACACACCACC 19089
QY 61	tactctcgctctttcatctatctcccggccaaggccctcatccccagccggcgcgagagat 120
DG 19090	TGCTTCCTTCGCTTTCATCATTCGCGGACAGGGCCTCATCCCCAGCGCGCGAGGAT 19149
QY 121	cgtgtgcacgtgctgtctctctgtggcgctggggctcatccttcttgggccaagtcc 180
DG 19150	CGTGTGCAGTGCCTGCTCTCTGCGGCGCTCGGGCTCATCTCTCTTGGGCCAGCTC 19209
QY 181	ttcagccagcaccccgacgttcttacgtatggagcccggtggcatgtgtgagaccac 240
DG 19210	TTCAAGCACACCCCGAGCTTCTACTGATGAGCCCGGTGGCATGTGTGAACACC 19269
QY 241	ctgtccaggaagcagcgggcaacctgcacattggccgtgcgcacctgatgcctctatc 300
DG 19270	CTGTGCGAGGGCAGCGGGCAACCTGCATGCGCTGGCGACCTGATCGGCTCTATC 19329
QY 301	tttttgtcacattgacgtggttgtagcttaaatgccacagagccgaaacctgtccgcc 360
DG 19330	TTTTTGTCCGACATGGACGTGTTGATGCTTACATGCCACAGAGCGGAAACCTGTCCGC 19389
QY 361	tttttaactgggaacagcgcgcgcgtgtgtctgcgcgcgcctgcagccacttccc 420
DG 19390	TTTTTCAACTGGGCAACAGACCCGCGCTGTCTCGCCGCCGCCCTGACAGCGCTTTCCC 19449
QY 421	cgaggcacctaatgcaagcaggcagctgcaagaacactgtcacgcggcgagccattcac 480
DG 19450	CGAGGCACCATCAGCAAGCAGGACGATGACAGACACTGTGCACGGCGAGCCCATTCAG 19509
QY 481	ctggcccgaggagcctgcgcctctaacagcaagtgtgtctaaggagtgcgcttcttc 540
DG 19510	CTGGCCCGGAGGCGCTGCGCTCTTACAGCCAGCTGTGCTAAGGAGGTGGCTTCTTC 19569
QY 541	aacctgcaggtgctctaccgcgtgctcagcaccocgcgctcaacctggcattcgtgac 600
DG 19570	AACCTGCAGTGTCTTACCCTGTCTCAGGACCCCGCCTCAACTCGCGCATCTGTGAC 19629
QY 601	ctggtgcgaccccgccggccgtgtgtgcctcccggaagcggcgggccgatactgaca 660
DG 19630	CTGGTGCAGCACCGCGGCGCTGTCTGCGCTCCCGGAGCGCGGCCGCTGATCTGCA 19689
QY 661	cgagcaaacgcatcgtgtgtggcaaccaacgcaagtgggtgagggcgacctcacctg 720
DG 19690	CGGCAACAGGATGTGTGTGGGCACCAACGCCAAGTGGGTGGAGCGGACCCCTACCTG 19749

repeat_region	/rpt_type=dispersed complement(462..771) /rpt_family="Alu"	repeat_region	/rpt_family="Alu" complement(15292..15589) /rpt_type=dispersed
repeat_region	772..945 /rpt_family="L2" /rpt_type=dispersed	repeat_region	15668..15718 /rpt_family="Alu" /rpt_type=dispersed
repeat_region	1169..1338 /rpt_family="MIR" /rpt_type=dispersed	repeat_region	15977..16098 /rpt_family="MIR" /rpt_type=dispersed
repeat_region	complement(1553..1804) /rpt_family="Alu" /rpt_type=dispersed	repeat_region	complement(17298..17433) /rpt_family="MIR" /rpt_type=dispersed
repeat_region	2616..2916 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	complement(18001..18310) /rpt_family="Alu" /rpt_type=dispersed
misc_feature	3309..3532 /note="similar to EST H12297 (EST cluster Hs.31147)"	repeat_region	complement(18420..18591) /rpt_family="MIR" /rpt_type=dispersed
misc_feature	3865..4028 /note="similar to EST H12297 (EST cluster Hs.31147)"	repeat_region	18732..18767 /rpt_family="Simple_repeat" /rpt_type=tandem
repeat_region	4099..4406 /rpt_family="Alu" /rpt_type=dispersed	CDS	18967..20202 /gene="CHST5"
repeat_region	4793..5250 /rpt_family="L2" /rpt_type=dispersed		/note="I-GlcNAc6ST: carbohydrate sulfotransferase 5; expressed in the small intestine and colon"
repeat_region	complement(5251..5329) /rpt_family="MIR" /rpt_type=dispersed		/product="intestinal N-acetylglucosamine-6-O-sulfotransferase"
repeat_region	complement(5768..6080) /rpt_family="Alu" /rpt_type=dispersed		/db_xref="GI:11023148"
repeat_region	complement(6081..6265) /rpt_family="Alu" /rpt_type=dispersed		/translation="MGMARVPKVAHSTRPAPARMPLRFSSKTVLVLLAOTTCLL LFLISRPSPSPAGGEDRVHVLVLSWRSGSSSLGOLFSDPDVFLMEPAHWVTWL SQGSAAFLHMAVRLMRSIFLCMDVDFDAPMQSRNLSAFENWATSEALQSPACSAF PRGTISKQDVCKTICTQPFSLAREACKSYSHVKEVFNLCVLLPLSDPALNLR IVHLVRDPAVLRSREAAAGPTLARDNGIVLTNGKWEADPHLEKLEVECKSHVRIAE AATLKPPPLGRVRLVRFEDLAREPLAEIRALYAFGLTLPQLEAWIHNTHTSGSI GKPLEAEPHTSRNARNVSOAWRHALPTKILRVQEVCAQALLGIRPVYSADQQRDL ILDLVLPGRGDHPSWASPD"
misc_feature	6111..6567 /note="similar to EST AA001593 (EST cluster Hs.6853)"	repeat_region	20504..20806 /rpt_family="Alu"
repeat_region	8035..8350 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	20869..20961 /rpt_family="L1"
misc_feature	10545..11049 /note="similar to EST AI950923 (EST cluster Hs.6853)"	repeat_region	20962..21200 /rpt_family="MER4-group"
repeat_region	11043..11195 /rpt_family="Alu" /rpt_type=dispersed	repeat_region	21325..21429 /rpt_family="MER4-group"
repeat_region	11200..11222 /rpt_family="Simple_repeat"	repeat_region	21430..21722 /rpt_type=dispersed
repeat_region	11570..11871 /rpt_family="Alu"	repeat_region	21725..22007 /rpt_type=dispersed
repeat_region	complement(12065..12198) /rpt_family="L1"	repeat_region	22008..22040 /rpt_family="Simple_repeat"
repeat_region	complement(12205..12493) /rpt_family="Alu"	repeat_region	22041..22313 /rpt_family="MER4-group"
repeat_region	complement(12813..13110) /rpt_family="Alu"	repeat_region	22325..22593 /rpt_type=dispersed
mRNA	join(14185..14228,16169..16263,17712..20821) /gene="CHST5"	repeat_region	22594..22949 /rpt_type=dispersed
gene	14185..20821 /product="intestinal N-acetylglucosamine 6-O-sulfotransferase"	repeat_region	23116..23221 /rpt_family="Simple_repeat"
repeat_region	14681..14713 /rpt_family="Simple_repeat"	repeat_region	complement(14995..15288) /rpt_type=tandem
repeat_region	complement(14715..14990) /rpt_family="Alu"		
repeat_region			

Query Match	100.0%;	Score 1173;	DB 9;	Length 3278;
Best Local Similarity	100.0%;	Pred. No. 2.5e-147;		
Matches 1173;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
QY	1	atgtggctgcacggttctccagcaagacagtgacagtgctctcctcggcacagaccac	60	
Db	1460	ATGTGGCTGCCACGGTCTCCAGCAAGACAGTAGACAGTGCTCTCTGGCCAGACACC	1519	
QY	61	tgcctcctgctcttcattcctcccgccagggccctcatccccagcgcgcgcgagagat	120	
Db	1520	TGCTCTCTGCTCTCATCATCTCCGGCCACGGGCCCTCATCCCCAGCGCGCGAGGAT	1579	
QY	121	cgtgtgcacgtgctgggtcctcgttgctgcctcgggtccatccctcttgggccagctc	180	
Db	1580	CGGTGCACGTGCTGGTGCTCTCTGGTGGCGCTCGGGCTCATCTCTTGGGCCAGCTC	1639	
QY	181	ttcagccagcaccccgacgtctcttaccctgatggagccgcgtggcatgtgtggaccac	240	
Db	1640	TTCAGCCAGCACCCCGACGCTCTTCTACCTCATGGAGCCCGCTGGCATGTGTGACCA	1699	
QY	241	ctgtcgcagggcagc	300	
Db	1700	CTGTCCAGGGCAGCGCGGCAACGCTGCGCATGCGCGTGGCGACCTGATGCGCTCTATC	1759	
QY	301	ttttgtgcacatgaacgtgtttgatgectacatgcacacagcgcgaacotgtgcgcc	360	
Db	1760	TTTTGTGGCAATGGACCTGTTTGATGCTACATGCCACAGAGCGGAAACCTGTCCGCC	1819	
QY	361	tttttcaactgggaacagc	420	
Db	1820	TTTTTCACTGGGCACAGCGCGCGCTGTGCTCGCCGCCGCCCTGACAGCGCTTTC	1879	
QY	421	cgaggcacatcagcaagcagcagctatgcaagacactgtgcacgcgcgcagccattcag	480	
Db	1880	CGAGGCACCATCAGCAAGCAGGACGATGCAAGACACTGTGCACGGCGGAGCAATCAG	1939	
QY	481	ctggcccgagcgctgcgcctctacacccagctgtgtcctaaggaggtgcgctctctc	540	
Db	1940	CTGGCCCGGAGGCGTGGCGCTCTTACAGCCAGTGGTGCTCAAGGAGTGGCTCTTTC	1999	
QY	541	aacctgcaggtgctctaccgcgtgcttcagcagcccgcgctcaacctgcgcctcgtgac	600	
Db	2000	AACTTGCAGTGTCTATCCCGTGTCTACGCGACCGCGGCTCAACTCGGCATCTGTCAC	2059	
QY	601	ctgggtgcgcacccgcggcgctgctgcctcccgaggcgcgcgccgatctggca	660	
Db	2060	CTGGTGCAGACCCCGCGCGCTGTGCGCTCCCGGAGCGCGCGGCCGATACCTGCA	2119	

Db	1004	CGCGAACCGGATCGTGTGGGACCAACGCGAAGTGGTGGAGCGCGCCCTCACCTG	1063
QY	721	cgctgattgcgaggtgtgcccgcagccactgcgcgtatccgcaggtccgcacactcaag	780
Db	1064	CGCTGATTCCGAGGTTGTCGCGCAGCCACGTCGCGCATCGCAGGCGGCACACTCAAG	1123
QY	781	cgccacacctctgcgcgcgcgtaccgcctggtgcgttcgaagaccitggcgcggag	840
Db	1124	CGCGACCCCTTCCTGCGCGCGCGCTACCGCTGCTGCGCTTCGAGGACCTGGCGGGAG	1183
QY	841	cgctggcagagatccgcgcactctacgccttcaccggcgtgacctcaagccacagctc	900
Db	1184	CCGCTGCACAGATCCCGGCACCTCTACGCTTCACCGGCGCTGACCTTCAGCCACAGCTC	1243
QY	901	gaggcctggatccacaacatcaaccacgcgggtcggggagtcggcaagccaatcgagccctc	960
Db	1244	GAGGCCCTGGATTCCACAACATCACCCACGGGTGCGGGGATCGCAAGCCAAATCGAGGCTTC	1303
QY	961	catacttcgtctagaaatggcgcaacgctcccgagcctcccgccctgagccagcgttgccctc	1020
Db	1304	CAIACCTTCGCTAGGAATGCGCGCAAGCTCTCCGACGCTGCGCGCCACGCGTTGCCCTTC	1363
QY	1021	actaagatcctgcgtgcagggaggtgtgcccgcgcgcgtgcagctgcgtgggtacacgg	1080
Db	1364	ACTAAGATTCCTGCGGTGCAGGAGGTGTGCGCGCGCGCTGCAGCTGCTGGGTACCGG	1423
QY	1081	octgtactctgcgcgaccagcagctgacccctaccctcgatctggtgctccacagggc	1140
Db	1424	CCTGTGTACTCTGCGACACAGCGTGACCTACCCCTGGATCTGGTGCTGCCACGAGGC	1483
QY	1141	ccagaccacttcagctgggcatcgccctgaactga	1173
Db	1484	CCAGACCACCTTCAGCTGGGCATCGCTGACTGA	1516
RESULT	3		
AF246718	3278 bp	mrna	linear
LOCUS	AF246718	Homo sapiens intestinal GlnAc-6-sulfotransferase (CHST5); mRNA,	PRI 31-OCT-2000
DEFINITION	complete cds, alternatively spliced.		
ACCESSION	AF246718		
VERSION	AF246718.1	GI:11055254	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
REFERENCE	1 (bases 1 to 3278)		
AUTHORS	Okamura,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A., and Fukuda,M.N.		
TITLE	Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulfotransferase gene		
JOURNAL	Nat. Genet. 26 (2), 237-241 (2000)		
MEDLINE	20472330		
PUBMED	11017086		
REFERENCE	2 (bases 1 to 3278)		
AUTHORS	Okamura,T.O., Nishida,K., Nakayama,J., Watanabe,H., Ozaki,K., Nakamura,T., Dota,A., Kawasaki,S., Inoue,Y., Maeda,N., Yamamoto,S., Fujiwara,T., Thonar,E.J., Shimomura,Y., Kinoshita,S., Tanigami,A., and Fukuda,M.N.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAR-2000) Glycobiology Program, The Burnham Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES	Location/Qualifiers		
source	1. .3278		
gene	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/chromosome="16"		
	/map="16q22: between D16S12 and D16S15"		
	1. .3278		
	/gene="CHST5"		

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 14:16:59 ; Search time 15810.8 Seconds
(without alignments)
1552.539 Million cell updates/sec

Title: US-09-593-828-4

Perfect score: 1173

Sequence: 1 atgtgggtgcacggttctc.....gctgggcgcctgactga 1173

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum Match 100%

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba:**

2: gb_htg:**

3: gb_in:**

4: gb_om:**

5: gb_ov:**

6: gb_pat:**

7: gb_ph:**

8: gb_pi:**

9: gb_pr:**

10: gb_ro:**

11: gb_sts:**

12: gb_sy:**

13: gb_un:**

14: gb_vi:**

15: em_ba:**

16: em_fun:**

17: em_hum:**

18: em_in:**

19: em_mu:**

20: em_om:**

21: em_or:**

22: em_ov:**

23: em_pat:**

24: em_ph:**

25: em_pl:**

26: em_ro:**

27: em_sts:**

28: em_un:**

29: em_vi:**

30: em_htg_hum:**

31: em_htg_inv:**

32: em_htg_other:**

33: em_htg_inv:**

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB	ID	Description

1	1173	100.0	1462	9	AF176839	Homo sapi
2	1173	100.0	2170	9	AF176838	Homo sapi
3	1173	100.0	3278	9	AF246718	Homo sapi
4	1173	100.0	71503	9	AF219991	Homo sapi
5	1173	100.0	157337	9	AC009163	Homo sapi
6	1173	100.0	159072	2	AC026419	Homo sapi
7	1173	100.0	194832	2	AC025287	Homo sapi
8	924.6	78.8	1647	6	AX327330	Sequence
9	924.6	78.8	2544	9	AF219990	Homo sapi
10	924.6	78.8	3786	9	AF280086	Homo sapi
11	924.6	78.8	159072	2	AC026419	Homo sapi
12	924.6	78.8	208185	2	AC009105	Homo sapi
13	745.2	63.5	1740	10	AF176841	Mus muscu
14	745.2	63.5	1989	10	AF176840	Mus muscu
15	711.6	60.7	139252	2	AC095664	Rattus no
16	400	34.1	2032	9	AF131235	Homo sapi
17	398.4	34.0	1333	9	AF149783	Homo sapi
18	398.4	34.0	1992	9	AF280088	Homo sapi
19	398.4	34.0	2011	9	AK026635	Homo sapi
20	398.4	34.0	183228	9	AC010547	Homo sapi
21	358.8	30.6	2201	10	AF109155	Mus muscu
22	354	30.2	1926	10	AF131236	Mus muscu
23	235.2	20.1	2156	6	AR071396	Sequence
24	235.2	20.1	2731	9	AB012192	Homo sapi
25	235.2	20.1	2156	6	EL4937	Human mRNA
26	235.2	20.1	6961	9	AB017915	Homo sapi
27	235.2	20.1	178251	9	AC073370	Homo sapi
28	235.2	20.1	196465	9	AC022392	Homo sapi
29	196.8	16.8	2354	5	CHK6ST	Chicken mRN
30	196.8	16.8	2354	6	AR050577	Sequence
31	196.8	16.8	2354	6	EL2369	DNA encodin
32	196.2	16.7	90256	9	AF001582	Homo sapi
33	196.2	16.7	137499	2	AC015931	Homo sapi
34	196.2	16.7	174187	2	AP002792	Homo sapi
35	196.2	16.7	215647	2	AC068591	Homo sapi
36	190.6	16.2	1425	10	AF178689	Rattus do
37	181	15.4	1654	10	AB008937	Mus muscu
38	181	15.4	7515	10	AB062107	Mus muscu
39	153.8	13.1	2227	9	AF083066	Homo sapi
40	153.8	13.1	3029	9	AK023268	Homo sapi
41	153.8	13.1	3535	9	AB021124	Homo sapi
42	153.8	13.1	4862	9	AB021125	Homo sapi
43	153.8	13.1	168056	2	AC055737	Homo sapi
44	153.8	13.1	191644	2	AC038450	Homo sapi
45	148.8	12.7	2393	6	AX321227	Sequence

ALIGNMENTS

RESULT 1
AF176839
LOCUS
DEFINITION Homo sapiens intestine N-acetylglucosamine 6-O-sulfotransferase (I-GlcNAc-6-ST) gene, complete cds.
ACCESSION AF176839
VERSION AF176839.1
KEYWORDS GI:5917707
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1462)

Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
Cloning and characterization of a mammalian N-acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue
Biochem. Biophys. Res. Commun. 263 (2), 543-549 (1999)

99423499
2 (bases 1 to 1462)
Lee, J.-K., Bhakta, S., Rosen, S.D. and Hemmerich, S.
Direct Submission
Submitted (10-AUG-1999) Respiratory Diseases, Roche Bioscience, JOURNAL

QY 121 cgtgtgcacgtgcctgtgctgtctctctgtgagcctcaggtcctcttctttggccagctc 180

[illegible]

QY	497	gcgctcctaagccacgttgctcaaggaggtggcgctcttcttaaacctgaagtgcct	556
Db	426	GCGACACTACAGCACGCTGGTGTGAATCCGTCGCCCTTCTTCGAGCTGCAGTCCCTGT	367
QY	557	accgcctgctcaagcaccccgcctcaacctggcatctgcaactgattgcgagaccgc	616
Db	366	ATCCGCTCCTTGACGAGACCCCAACTGAACCTCCGCATCGCTCCACCTGGTCCGGACCCCTC	307
QY	617	gggcctgtgctgcgtccccggagcggcgcccgatctggtgcaacgcgacaacgcgatcg	676
Db	306	GGCGCGTGGCCGCTCCAGAGAACAAGTCGGCCAACGCTTCGTGCGCGACGACGCCGTCA	247
QY	677	tgtgtggcacacaacggcaagtgggtgagggccacctcaacctgcgccttatctgcaggg	736
Db	246	TCTTTGGAACAACAGCGCA---TGGCGCGCGCGAGGTGCTGTATPACGCTCATGACAGAGA	190
QY	737	tgtgcgcagccacgtgcgcatacgcgagcgcgccacactcaacgcccacccttcctgc	796
Db	189	TC TGCGCAGCCACGTGGCATACAGCAGAGGGCCAGCTGAAGCCCGCCTTTCTCA	130
QY	797	gcggcgcctaccgctggtgcttcagagacctggcgggagccgctggcagagatcc	856
Db	129	AAGGCGCTACAAAATGGTCCGCTTCGAGGACCTGTGTGCGGACCCTGTCGGGAGATCG	70
QY	857	gcgaacttaacgcttacacgcgcctgcacctcaacgcacacagctcagcctgtagatcacaa	916
Db	69	GCGGCATCTCAAGATTCTGGTCTGGAGATGACCAAGCAGCTGGCGCAGTGGATTACC	10
QY	917	acatacc 924	
Db	9	AGCTTCC 2	
RESULT	8		
AG035205		AG035205	657 bp DNA linear GSS 01-NOV-2001
LOCUS		Pan troglodytes DNA, clone: PTB-010M13.F, genomic survey sequence.	
DEFINITION		AG035205	
ACCESSION		AG035205.1 GI:16562078	
VERSION		GSS; GSS (genome survey sequence).	
KEYWORDS		Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male	
SOURCE		BAC Library clone:PTB-010M13.F.	
ORGANISM		Pan troglodytes	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.	
TITLE		1 (sites)	
JOURNAL		Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,	
REFERENCE		Totoki,Y., Watanabe,H. and Sakaki,Y.	
AUTHORS		BAC end sequences of Library PTB	
TITLE		Unpublished	
JOURNAL		2 (bases 1 to 657)	
REFERENCE		Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,	
AUTHORS		Totoki,Y., Watanabe,H. and Sakaki,Y.	
TITLE		Direct Submission	
JOURNAL		Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical	
REFERENCE		and Chemical Research (RIKEN), Genomic Sciences Center (GSC);	
AUTHORS		1-7-2 Suehiro-chou,Isumi-ku, Yokohama, Kanagawa 230-0045, Japan	
TITLE		(E-mail:chumpbes@sc.riken.go.jp, URL:http://hgpc.gsc.riken.go.jp/;	
JOURNAL		Tel:81-45-503-9111, Fax:81-45-503-9170)	
REFERENCE		Clones are derived from the chimpanzee BAC library PTB This BAC end	
AUTHORS		was generated during the R&D process and may have higher chance of	
TITLE		clone tracking errors.	
JOURNAL		PRIMERS	
REFERENCE		Sequencing: -21M13	
AUTHORS		LIBRARY	
TITLE		Vector : pKS145	
JOURNAL		R.Site 1 : SacI	
REFERENCE		R.Site 2 : SacI	
AUTHORS		Location/Qualifiers	
TITLE		1. .657	
JOURNAL		/organism="Pan troglodytes"	
REFERENCE		/db_xref="taxon:9598"	
AUTHORS		source	

BASE COUNT		85 a		237 c		201 g		133 t		1 others	
ORIGIN											
/clone="PTB-010M13.F"											
/sex="male"											
/cell_type="lymphoblast"											
/clone_lib="PTB Chimpanzee Male BAC Library"											
Query Match 26.5%; Score 310.6; DB 12; Length 657;											
Best Local Similarity 86.2%; Pred. No. 2.5e-44;											
Matches 368; Conservative 0; Mismatches 54; Indels 5; Gaps 2;											
Qy	1	atgaggctgcacaggttctccagaaagacagtgcagctgtctctctctggtgcacacacc	60								
Db	234	avtgggtgcgcggcggtctccagacacacacacacacacacacacacacacacacac	290								
Qy	61	tgcctctgtcttcttcacatctccgggcagggcccttcacccacagcggcgaggat	120								
Db	291	ttgcctctccctctcttctctgtttccgggcagggccctgcgtccacagcggcgagg	350								
Qy	121	cgatgacagctgctgctctctctctggtgcgtgcggctcactctcttggccaactc	180								
Db	351	cggtgtcagatgctggtctctctctggtgcgtgcggctcactctcttggccaactc	410								
Qy	181	ttcagcgcagcaccgcagcttcttacctgatgagccgcgtgtgcattgtgtgacacc	240								
Db	411	ttcaacacagcaccgcagcttcttacctgatgagccgcgtgtgcattgtgtgacacc	470								
Qy	241	ctgtgcagggcagcgcggcgaacgtgcacatggcgtgcgcgacctgatcgctctatc	300								
Db	471	ctgtgcagggcagcgcggcgaacgtgcacatggcgtgcgcgacctgatcgctctatc	530								
Qy	301	tttttgtcgaca--tggaactgtttgatgctacatgccagagccgaacacctgtcg	358								
Db	531	ttcctgttcggacacgtgcagctgtttgatgctacatgccagagccgaacacctgtcg	590								
Qy	359	ctttttcaactgggcaacagcgcgcgtgtgtctgcgcgcgcgcgcgcgccttc	418								
Db	591	acctcttcagttgggcccgtgagccgcgtgcacatgctgcgcgcgcgcgcgccttc	650								
Qy	419	cccgagg 425									
Db	651	cccgagg 657									
RESULT 9											
A1115260											
LOCUS	A1115260 568 bp mRNA linear EST 02-SEP-1998										
DEFINITION	IMAGE185164.1 Sugano mouse embryo newa Mus musculus cdna clone										
	6-SULFOTRANSFERASE ; mRNA sequence.										
ACCESSION	A1115260										
VERSION	A1115260.1 GI:3515584										
KEYWORDS	EST.										
SOURCE	house mouse.										
ORGANISM	Mus musculus										
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.										
AUTHORS	Marra, M., Hallier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellengberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.										
TITLE	The WashU-HMI Mouse EST Project										
JOURNAL	Unpublished (1996)										
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of MedicineP 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu										

[illegible]

QY	603	ggtgcgcagcccgccggccgtgtctgcctcccggagagcgccggcccgatactatcgccacg	662
Db	196	GGTGGCGGACCCGGCGGCGCGTGTCTGTCGCTCCGGGAGGCGGCGGCGCGTACTGGCAGC	137
QY	663	cgacaacgcatctgtctgggcacacaacgcaagtggtagagccagccacctaccctgcg	722
Db	136	CGACAACGGCATCTGTCTGGGCGACCAACGGCAAGTGGTGGAGGCGCCACCCCTACCTGGCG	77
QY	723	ccctgattcgaggtgtgcgcagccacgtctgcctgcgcagagcgccgacacactcaagcc	782
Db	76	CCTGATTCGGAGTGTGCGCAGCCACGTCGGCATGCCGAGGCGGCCACACATCAGCC	17
QY	783	gccacccttcctgcgc	798
Db	16	GCCACCTTCCTGCGC	1

RESULT	6
LOCUS	AK009113
DEFINITION	Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003G18;carboxylate (chondroitin 6/keratan) sulfotransferase 4, full insert sequence.
ACCESSION	AK009113
VERSION	AK009113.1 GI:12843701
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (strain:C57BL/6J) adult male tongue cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library clone:2310003G18.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Carninci,P. and Hayashizaki,Y.
TITLE	High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2 (sites)
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20499374
PUBMED	11042159
REFERENCE	3 (sites)
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagao,K., Sasaki,N., Carninci,P., Konno,H., Akiyama,Y., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakauchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4 (sites)
AUTHORS	FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5 (bases 1 to 1923)
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 917 Std Error: 0.00
Seq primer: --40UP from Gibco
High quality sequence stop: 490.

[illegible]

```

1.620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2405856"
/clone_lib="NCI_CGAP_lu19"
/tissue_type="squamous cell carcinoma, poorly
differentiated (4 pooled tumors, including primary and
metastatic)"
/dev_stage="adult"
/lab_host="PH10B (phage-resistant)"
/notes="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
pooled lung tumor tissue, and was then primed with a Not I
- cligo(dT) primer. Double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization. Library constructed by Bento Soares and M.
Fatima Bonaído."
106 a 195 c 227 g 89 t 3 others

```

Query Match 41.6%; Score 487.6; DB 9; Length 620;
Best Local Similarity 87.4%; Pred. No. 7.1e-75;
Matches 543; Conservative 0; Mismatches 77; Indels 1; Gaps 1;

184	QY	agcagcaccocagactcttcttaactgataagaccgcgtggcatgtgtgacacacctg	243
620	Db	AACCAGACCCGAGCGGTCTTCTACNTAATGAGCGCGGTGGCNACGTTGGACCA-CCTG	562
244	QY	tgcagggcagcgcgcgaacgcctgcacatggcgtggcgacactgatgcgctctatcttt	303
561	Db	TGCAGGGCAGCGCGCAACGCCTGCACATGGCTGTCGGACCTGTGGCGCTCGCTCTTC	502
304	QY	ttdtgacactggacgtgtttgatgcctacatgccacagcgcgaacctgtccgcttt	363
501	Db	CTGTGGACATGACGCTGTTTGATGCTATCTGCTTGGCGCGCAACCTGCCGACCTC	442
364	QY	tccaactgggcaacgagccgcgcgtgtgtctgcgcgcgcgccttcacccga	423
441	Db	TTCCAGTGGGCGGTGAGCGTGACATGTGTCGCCACCGCGCTGCAGTGCCTTGC	382
424	QY	ggcaccatcagcaagcaggacgtatgcaagacactgtcacgcgcgcagccattcacctg	483
381	Db	GGCGCCATCAGCAGCGAGGCGGTGTGCAGGCCACTGTGCGCGCGCAGTCTTTCACCC	322
484	QY	gcocggaggcctgcgcctctctacagccacgtgtgtctcaaggaggtgcgcttctcaac	543
321	Db	GCCGGGAGGCGTCCGCTCTACAGCCACGTGTGCTCAAGGAGTGGCTTCACAC	262
544	QY	ctcaggtactctaccgcctgtctcagcaccgcgcctcaacctgcacatgtgcacctg	603
261	Db	CTCAGAGTGTCTTACCCGCTGTCTCAGCGACCGCGCTCAACCTACGCATCGTGCACCTG	202
604	QY	gtgcgcgaccgcggcgctgtcgcctccggagagcggcgccgatactggcagcgc	563
201	Db	GTCCGGCACCCGGCGCGGTGCTGCGCTCCCGGGACGACAGCGCTCTGGCGGT	142
664	QY	gacacggcatcgtgtcgtggcaccacgcgaagtgggtggagccgcacctcaactggc	723
141	Db	GACACGGGATCTGTGTGGGCACCAACGGCAGCTGGGTGGAGCCGACCCCGGCTTGCC	82
724	QY	ctgattcggagggtgtgcgcgacacacgtgcgacatgcgcgagcgccacatcaagcgc	783
81	Db	GTGGTCCGGGAGGTGTGCGGTATGCCACGTACGCATCGCAGAGCGCGCCACATCTAAGCG	22

QY 784 ccacccttctgcgggcgc 804
||||||| |||||||
Db 21 CCACCCCTTTNTGCGGGCGC 1

RESULT 5

AW081348	linear	436 bp	mrna	EST 14-OCT-1999
xc41d06.x1	NCI-CGAP Co20 Homo sapiens cDNA clone IMAGE:2586803	3'		
DEFINITION	similar to TR:068199 068199 CHONDROITIN 6-SULFOTRANSFERASE. ; mRNA sequence.			
ACCESSION	AW081348			
VERSION	AW081348.1	GI:6036500		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 436)			
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .			
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index			
COMMENT	Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgaps-r@mail.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html Seq primer: -400P from Gibco High quality sequence stop: 431.			

	source
FEATURES	

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="586803"
/clone_lib="NCI_CGAP_Co20"
/tissue_type="moderately differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-spOrf6; Site_1: Sali;
Site_2: NotI; Cloned unidirectionally. Primer: oligo dr.
Normalized to Cot 500. Average insert size 1.1kb.
Normalized version of NCI_CGAP_Col8. Library constructed
by Life Technologies."
64 a 136 c 169 g 67 t

```

BASE COUNT
ORIGIN

```
Query Match          35.9%; Score 420.8; DB 9; Length 436;
Best Local Similarity 99.3%; Pred. No. 2.3e-63;
Matches 433; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

364	QY	ttcaactgggcaacgagcgcgcgtgtgtctgcgcgccgcctgcagcgcctttcccca	423
436	Db	TTCAACTGGGACAGACCGCGGCTGTGTGTGCCGCCCGCTGCMGCGCCTTTCGCCGA	377
424	QY	ggcaccatcagcaagcaggacgtatgcaagacactgtgcacgcgcgcacccattcagctg	483
376	Db	GGCACCATCAGCAAGCAGACGCTATGTCAAGACACTGTGCACGCGCAGCATTCAGCCTG	317
484	QY	gcccggagggcctgcgcctcttcacgccaagtgggtgctcaaggagggtgcgcttcttcaac	543
316	Db	GCCCGGAGGGCTGCGCGTCTTACAGCCACGTGGTGTCTCAAGGAGGTGGCTTCCTCCAC	257
544	QY	ctgcagg-tgcctaccgcctgctcagcagaccgcgcgtccaacctgcgaatcgtgcacct	602
256	Db	CTGCAAGTTGCTCTACCCGCTGCTCAGCGACCCGGCTCAACCTGGGATCGTGACCT	197

QY 443 acgtatgcaagacactgtgcaagcgagcgagcattcagcctggccgggagggcctgcgct 502
 Db 376 AGGTATGCAAGACACTGTGCAAGCGGCGAGCCATTCAAGCTGGCCGGAGGCGCTGCGGCT 317
 QY 503 cctacagccacgtgtgtctcaaggaggtgcgtcttcttaacactgcagggtgcttaaccgc 562
 Db 316 CCTACAGCCACGCTGGTGTCTCAAGAGGTGCGCTTCTTAACCTGCAAGGTGCTTACCCGC 257
 QY 563 tgcctagcagcccgctgcaactgcgactgcgactgcgactgcgactgcgactgcgactgcg 622
 Db 256 TGCTAGCAGCCCGCGCTCAAGCTGCGATCGTGCACCTGGTGGCGAGCCCGGCGCG 197
 QY 623 tgcctgcgtcccgaggagcgcgcgccgactgcgactgcgactgcgactgcgactgcgactgcg 682
 Db 196 TGCTGGCGCTCCCGGAGGCGCGCGCCCGATGCTGCGCGGCGACACGGGATCGTGTGG 137
 QY 683 gcaccacggcgaagtgggtggaggccgacccctcaactgcgactgcgactgcgactgcgactgcg 742
 Db 136 GCACCAACGCAAGTGGGTGGAGCGCGACCTCACCTGGCGCCCTGATTTGGAGGTGTTC 77
 QY 743 gcagccactgcgactgcgagcgcgccacactcaagccgacaccccttctcgcgcggcc 802
 Db 76 GCACCAACGCTGGGATGGCGAGCAACACACTCAAAACCCGCCACCTTTTGGGGGGCG 17
 QY 803 gcta 806
 Db 16 GNCA 13
 RESULT 3
 BF197521/c
 LOCUS 525 bp mRNA linear EST 03-NOV-2000
 DEFINITION 7084a08.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'
 similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 ACCESSION BF197521
 VERSION BF197521.1 GI:11086670
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 525)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapb-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
 High quality sequence stop: 451.
 Location/Qualifiers
 1..525
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /clone_lib="NCI_CGAP_Kid11"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pMT3b-Pac (Pharmacia) with
 a modified polylinker; Site:1; Not.1; Site.2; Eco RI;
 Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library

(cloneIDs 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo. "

BASE COUNT 90 a 163 c 187 g 85 t
 ORIGIN

Query Match 44.8%; Score 525; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2.4e-81;
 Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 226 catgtgtgaccacccctgtcgcaggcagcgcgcaacgctgcacatggccgtgcgcgac 285
 Db 525 CATGTGTGGACCAACCTGTGCGAGGCGAGCGCGCAACGCTGCACATGGCCGTGCGGAC 466
 QY 286 ctatgcgctctatcttttgcgacatggaactgtttgatgctcatatgcacagcagc 345
 Db 465 CTGATGGCGCTATCTTTTGTGGACATGAGACCTGTTTGTATGCTTACATGCGACAGC 406
 QY 346 cgaacactgtccgcttttcaactgggcaacgagcgcgctgtgtctgcgcgcgcgc 405
 Db 405 CGAAACCTGTCCGGCTTTTCAACTGGCAACGAGCGCGCTGTGTCTCGCCGCCGCC 346
 QY 406 tgaagcgctttcccgaggcaccatcagcaagcagcagcgtatcaagacactgcacg 465
 Db 345 TGCAGCGCCTTTCCTCCGAGGCGACCATCAGCAAGCAGGACGATGCAAGACACTGTGCAG 286
 QY 466 cggcagccattcagctggccgggagcctgcgcctctcacagccacgctgtgtcctcaag 525
 Db 285 CGGCAGCCATTCAGCCTGGCCGGAGGCGCTGCCCTCTTACAGCCACGCTGTGCTCAAG 226
 QY 526 gagtgcgctcttcaacctgcaggtgctctacccgctgtcgcagcccgcgctcaac 585
 Db 225 GAGTGGCGCTTCTTCAACCTGCAGTGTCTTACCGCTGCTCAGCGACCCCGCGCTCAAC 166
 QY 586 ctgcgactgcacactggtgcgacccgcgcgctgctgcctcccgagagcgcg 645
 Db 165 CTGCGCATCGTGACCTGGTGGCGGACCCCGCGCGCTGCTGCGCTCCCGGAGGCGCG 106
 QY 646 ggcgcgatactggcagcgcaacggcctgtgtggcaccacggcaagtgggtggag 705
 Db 105 GGCCCGATCTGCGACGCGACACCGCATCTGTGTGGSCACCAACGGCAAGTGGGTGGAG 46
 QY 706 gccacactcactgcgcctgattgcgaggtgagcgcgcgcgcgc 750
 Db 45 GCCGACCTTCACCTGGCGCTGATTGCGAGGTGTGCCGACGCCAC 1

RESULT 4
 AI824100/c

LOCUS AI824100 620 bp mRNA linear EST 21-DEC-1999
 DEFINITION WJ46c01.x1 NCI_CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2408856 3'
 similar to TR:075667 075667 DJ7116.4 ;contains PTRS.b2 PTRS
 repetitive element ;, mRNA sequence.

ACCESSION AI824100
 VERSION AI824100.1 GI:5444771

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 620)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapb-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 24, 2002, 09:23:53 ; Search time 10470.7 Seconds
(without alignments)
1512.025 Million cell updates/sec

Title: US-09-593-828-4
Perfect score: 1173
Sequence: 1 atgtgctgcacggtctc.....gctgggcgcctgactga 1173

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hnc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	657.4	56.0	695	10	BE857538	BE857538 7g01a08.x
C 2	631	53.8	735	10	BE858652	BE858652 7g01a09.x
C 3	525	44.8	525	10	BF197521	BF197521 7o84a08.x
C 4	487.6	41.6	620	9	AI824100	AI824100 w146c01.x
C 5	420.8	35.9	436	9	AW081348	AW081348 xc41b06.x
C 6	352.4	30.0	1923	11	AK009113	AK009113 Mus muscu
C 7	346.4	29.5	1067	12	CNS03KN7	AL248380 Tetraodon
C 8	310.6	26.5	657	12	AG035205	AG035205 Pan trogl
C 9	290.6	24.8	568	9	AI115260	AI115260 u143c07.y
C 10	284.2	24.2	954	10	BI823850	BI823850 603039012
C 11	278	23.7	849	12	CNS04QFN	AL302540 Tetraodon
C 12	271.2	23.1	517	10	BF042384	BF042384 BP250022A
C 13	260	22.2	260	10	BI824736	BI824736 603033615
C 14	240.2	20.5	433	12	AZ405100	AZ405100 LM0173108
C 15	239.8	20.4	429	9	BB849113	BB849113 BB849113
C 16	224.4	19.1	695	10	BG964671	BG964671 602831875
C 17	224.4	19.1	783	10	BG963298	BG963298 602827716

18	218.8	18.7	852	10	BG966340	BG966340 602832826
19	174.6	14.9	571	10	BM129080	BM129080 i17c04.y
20	159.4	13.6	965	10	BF579746	BF579746 602095056
21	157.6	13.4	599	10	BJ031352	BJ031352 BJ031352
22	148.6	12.7	322	9	AI156825	AI156825 u144c08.y
23	146.6	12.5	536	10	BM245312	BM245312 K0722H07-
24	134.6	11.5	2070	11	AK011202	AK011202 Mus muscu
25	134.4	11.5	793	10	BI102274	BI102274 602885587
26	131.8	11.2	497	10	BM246681	BM246681 K0741E04-
27	130.6	11.1	955	10	BG107354	BG107354 602290740
28	129.2	11.0	634	9	AU180328	AU180328 AJ180328
29	116.2	9.9	662	10	BF344303	BF344303 602017370
30	111.2	9.5	1070	10	BF163765	BF163765 601769868
31	109.2	9.3	540	9	AW112223	AW112223 ug47g02.y
32	103	8.8	103	9	AI282873	AI282873 qt87e06.x
33	103	8.8	297	9	AA261202	AA261202 va49a06.f
34	103	8.8	500	9	AI529474	AI529474 va49a06.y
C 35	99.8	8.5	135	9	AI824198	AI824198 wj36a01.x
36	97.4	8.3	494	9	AI088880	AI088880 qai7a07.x
37	97.4	8.3	539	10	BF593996	BF593996 nac20a08.
38	97.4	8.3	604	10	BE857485	BE857485 7f98g02.x
39	97.4	8.3	722	9	AW027448	AW027448 w73h09.x
40	96.6	8.2	714	10	BI115837	BI115837 602866294
41	96.4	8.2	675	9	AI939595	AI939595 tf88ell.x
42	96.2	8.2	564	10	BM488009	BM488009 pgm2h.pk0
43	94.8	8.1	828	10	BI665990	BI665990 603287043
44	94.2	8.0	468	12	AQ939450	AQ939450 NR1-153R
45	91	7.8	652	9	AL646758	AL646758 AL646758

ALIGNMENTS

RESULT 1
LOCUS BE857538/c
DEFINITION BE857538 695 bp mRNA linear EST 29-SEP-2000
7g01a08.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone IMAGE:3305174 3,
similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE. ;, mRNA sequence.

ACCESSION BE857538 GI:10371664

VERSION BE857538.1

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 695)

AUTHORS Fukuyota; Metazoa; Chordata; Vertebrata; Euteleostomi;

TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BrGAP), Tumor Gene Index

Unpublished (1998)

CONTACT: Robert Strausberg, Ph.D.

Email: cgaps-r@mail.nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL, send email to:

info@image.llnl.gov

Seq primer: -40UP from Gibco

High quality sequence stop: 431.

Location/Qualifiers

1..695

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3305174"

/clone_lib="NCI_CGAP_Brn23"

/tissue_type="glioblastoma (pooled)"

/lab_host="DB10B"

FEATURES
source

```

QY 30 PGPSSPAG---GEDRVHVLVLSWRSGSFLGQLFSQHPDVFYLMPEPAHWVWTTLSQGS 86
Db 85 PGNLSAVGEAVTQEKOHYVHATWRTGSSFLGELFNQHPDVFYLDPMWHLRQALYPGDA 144
QY 87 ATLHMVAVRDLMSIFLDCMDVFDAY-----MPSRNL--FFNWATSRALCSPPA 135
Db 145 ESLQALRDLMSLPRCDFSVLRVLAQPDGGERAPDSANLTATLFRWTKNKVICSPP 204
QY 136 CSAFERGT---ISKQDVCKTLCRQPESL-AREA-CRSYSHVVLKEVFFNLOVLYPLL 189
Db 205 CPAAPRARADVGLVEDKACESTC--PPVSLRALEAECKYPVVVIKDVRLDLGLVLP 262
QY 190 SDPALNLRIVHLVDPRVAVLRSREAAGPILARDNGIVLGTNGK-----WVEADP 238
Db 263 RDPGLNLKVVQFLFRDPRAVHNSRLKSRQGLLRRESIOVLTRQGRDHFHVRVLAHGV 322
QY 239 --HLRIREVCRSHVRIAEEA-----TLKPPFLRGYRLVRPEDIARPELA 283
Db 323 GGQARALPSAPRADFFLTSALEVCIAWLRLDLFTRGAPAWLRRYRLRYEDLVWQ 382
QY 284 EIRALYAFTGLTLPQLEAWIHNTTHSGGIGKPIEAFHTSSRNARNVSOQWRHALPFTKI 343
Db 383 QLRRLRFSGLRTLAALDAFANWTRGSAGAD-RPFHLSARDAREAVHAWRERLSQ 441
QY 344 LRQVEVCAGALQILGY 359
Db 442 RQVETACAPAMRLIAY 457

```

```

RESULT 15
ID 075667 PRELIMINARY; PRT; 486 AA.
AC 075667;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE DJ71L16.4 (N-ACETYLGUCOSAMINE-6-O-SULFOTRANSFERASE).
GN DJ71L16.4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Graham D.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Uchimura K., Muramatsu T.;
RT "Identification and molecular characterization of a cDNA encoding a
RL novel N-acetylglucosamine-6-O-sulfotransferase.";
DR EMBL; AL022165; CAAL18154.1;
DR EMBL; AB040711; BAB13770.1;
KW Transferase.
SQ SEQUENCE 486 AA; 54266 MW; 3F1FD1430B3C8E95 CRC64;

```

```

Query Match      29.7%; Score 610; DB 4; Length 486;
Best Local Similarity 38.8%; Pred. No. 4.7e-49;
Matches 147; Conservative 52; Mismatches 126; Indels 54; Gaps 11;

QY 30 PGPSSPAG---DRVHVLVLSWRSGSFLGQLFSQHPDVFYLMPEPAHWVWTTLSQGS 85
Db 86 PGNLSAVGEAVSRKQHIYVHATWRTGSSFLGELFNQHPDVFYLDPMWHLRQALYPGDA 145
QY 86 AATLHMVAVRDLMSIFLDCMDVFDAY-----MPSRNL--SAFFNWATSRALCSPP 134
Db 146 AESLQALRDLMSLPRCDFSVLRVLAQPDGGERAPDSANLTATLFRWTKNKVICSPP 205
QY 135 ACSAFPRGTISKQDVCKTLCRQ--PESL-AREA-CRSYSHVVLKEVFFNLOVLYPLL 190

```

```

Db 206 LCPGAPRAAEVGLVEDTACERSCPPVATRALEAECKYPVVVIKDVRLDLGLVLP 265
QY 191 DPALNLRIVHLVRDPRA-----VLSREAG-----PILARDNGIVLG 228
Db 266 DEGLNLKVVQFLFRDPRAVHNSRLKSRQGLLRRESIOVLTRQGRDHFHVRVLAHGV 325
QY 229 TNGKWEADPH-----LRLIREVCRSHVRIAEEAATLKPPFLRGYRLVRPEDIAR 280
Db 326 GQSRALPAAPRADFFLTGALVEICEAWLRLDLFARGA-----PAWLRRLRYEDLV 381
QY 281 PLAEIRALYAFTGLTLPQLEAWIHNTTHSGGIGKPIEAFHTSSRNARNVSOQWRHALP 340
Db 382 PRAQLRRLRFSGLRDLAALDAFANWTRGAAGAD-RPFHLSARDAREAVHAWRERLSR 440
QY 341 TKILRVOEVCAGALQILGY 359
Db 441 EQVRQVEAACAPAMRLIAY 459

```

Search completed: September 20, 2002, 20:02:26
Job time: 255 sec

Db	217	ARMDLSALRYCDLSVFQQLSPAGSGGRNLTTLGIFGAATNKVCSPLCPAYRKEVGLV	276
QY	148	D--VCKTCTCTROPFSLARACRSYSHVVLKVEYRNFNLOVYLPFLSDPALNLRIVHLVYDP	205
Db	277	DDRVCCK-CPPQRLARFECECKRYETLVKGVKVFDFVAVLAPLLEDDPALDLKVLHLVRDP	335
QY	206	RAVLS-----REAAGPILAD-----NGVLGTNGKV--EADPH-LRL	242
Db	336	RAVASRIRSRHGLIRESLOVRSRDPRAHRMPFLEAGHKLGAKEGVGGPADYHALGA	395
QY	243	IRVCRSHVRVIAEAATLKPPPLRGYRLVRFEEDLAREPLAEIRALYAFYGLTLPQLEA	302
Db	396	MEVICNSMAKTLQTA-LQPPDWLQGHYLVRYEDLVGDVKTLLRRVYDFVGLVSPVMEQ	454
QY	303	WINITHGSG-IGKPIEAFHTSSRNARNYSQARHAPFTKILRVQEVACAGALQLGLYRP	361
Db	455	FALNMTSGSGSSSKP---FVVSARNATAQANAWRTALTFFQIKQVEEFCYQPMVILGYER	511
QY	362	VYSADQQRDLTDLV	376
Db	512	VNSPEEVKLSKILL	526
RESULT	13		
Q9EP78	ID	Q9EP78	PRELIMINARY; PRT; 484 AA.
AC	Q9EP78;		
AD	Q9EP78;		
DT	01-MAR-2001 (TrEMBLrel. 16, Created)		
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE GST-5 (N-		
DE	ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE) (2600013M07RIK PROTEIN).		
DE	CHST7 OR GS75 OR 2600013M07RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J;		
RX	PubMed=10956661;		
RA	Bhakta S., Bartes A., Bowman K.G., Kao W.M., Polsky I., Lee J.-K.,		
RA	Cook B.N., Bruehl R.E., Rosen S.D., Bertozzi C.R., Hemmerich S.;		
RT	"Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2		
RT	(GST-5)".		
RL	J. Biol. Chem. 275:40226-40234(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Uchimura K., Muramatsu T.;		
RT	"Identification and molecular characterization of a cDNA encoding a		
RT	novel N-acetylglucosamine-6-O-sulfotransferase."		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=EMBRYO;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arai K., Harada A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,		
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.		
RA	Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming		

RL Genomics 55:345-347(1999).
 DR EMBL: AF083066; AD20981.1; --
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferase.
 SQ SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;

Query Match
 Best Local Similarity 32.2%; Score 659.5; DB 4; Length 530;
 Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;

QY 31 GPSSPAG-----GEDRVHVLVLSWRSGSFLGQLFSQHPDVFYLMPEPAHWVWTL 81
 Db 145 GVAAPGNGTRGGVGDKRLQVYVTTWRSGSFFGELFNQNPFFLYEPVHWVWQKL 204
 QY 82 SOGSAATLHMYVDLMRSIFLCMDVFDAYMPQ---SRNLS--AFFNWAISRALCSPAC 136
 Db 205 YPGDAVSGGAARDMLSALYRCDLSVFQYSPAGSGGRNLTLTGIFGAATNKVVCSSPLC 264
 QY 137 SAFFRGTTISKQD--VCKTLCTRPQSLAREACRSYSHVVLKEVRFNLOVLYPLSLDPAL 194
 Db 265 PAYREVVGLVDRVCKK--CPQRLARFEECKRYTLVIKGVRFDVAVLAPLLRDPAL 323
 QY 195 NLRVHLVDRPVLRS-----REAAGPILARD-----NGIVLTNGKWV 234
 Db 324 DLKVHLVDRPRAVASSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAGHKLGAKEGV 383
 QY 235 --EADPH--LRLREYCRSHVRIAETATLKPPFLRGYRLVRFPDLAREPLAEIRALYAF 291
 Db 384 GGPADYHALGAWETVCSMAKTLQTA--LQPPDWLQGHYLVVRYEDLVGDPVKTLRRVYDF 442
 QY 292 TGLTLTPOLEAHIHNTHGSG-IGKPIEAFHTSSNARNVSOAHRHALPFTKILRVQVRC 350
 Db 443 VGLVSPMEQFALNMTSGSSSKP---FVVSARNATQAAANWARTALTFOQIKQVVEFC 499
 QY 351 AGALQLLGYRPVYSADQOORLTLDLV 376
 Db 500 YQPMVILGYERNVSPVVKDLSTLL 525

RESULT 11
 Q9UED5 PRELIMINARY; PRT; 484 AA.
 ID Q9UED5
 AC Q9UED5
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE (GLCNAC6ST).
 GN GN6ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98391845; PubMed=9722682;
 RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
 RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
 RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
 RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
 RT mapping, and expression in various organs and tumor cells.";
 RL J. Biochem. 124:670-678(1998).
 DR EMBL: AB014679; BAA34265.1; --
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferase.
 SQ SEQUENCE 484 AA; 53641 MW; D077EBCD7645F2F0 CRC64;

Query Match
 Best Local Similarity 32.0%; Score 656.5; DB 4; Length 484;
 Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

Best Local Similarity 39.7%; Pred. No. 1.9e-53;
 Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGGEDRVH-VLVLSWRSGSFLGQLFSQHPDVFYLMPEPAHWVWTLTGSGSAATLHMA 92
 Db 110 APEGVGDKRHWMYVFTTWRSSTGSGSFFGELFNQNPFFLYEPVHWVWQKLYPGDAVSLQGA 169
 QY 93 VRDLMSRIFLCMDVFDAYMPQ---SRNLS--AFFNWAISRALCSPACSAFFRGTTISKQ 147
 Db 170 ARDLMSALYRCDLSVFQYSPAGSGGRNLTLTGIFGAATNKVVCSSPLCPAYRKREVVGLV 229
 QY 148 D--VCKTLCTRPQSLAREACRSYSHVVLKEVRFNLOVLYPLSLDPALNLRVHLVDRP 205
 Db 230 DDRVCKK--CPQRLARFEECKRYTLVIKGVRFDVAVLAPLLRDPALDLKVHLVDRP 288
 QY 206 RAVLRS-----REAAGPILARD-----NGIVLTNGKWV--EADPH--LRL 242
 Db 289 RAVASSRIRSRHGLIRESLQVVRSDPRAHRMPFLEAGHKLGAKEGVGPGADYHALGA 348
 QY 243 IREYCRSHVRIAETATLKPPFLRGYRLVRFPDLAREPLAEIRALYAFTLTLPQLA 302
 Db 349 MEVICNSMAKTLQTA--LQPPDWLQGHYLVVRYEDLVGDPVKTLRRYDFVGLVSPMEQ 407
 QY 303 WLNHTHGSG-IGKPIEAFHTSSNARNVSOAHRHALPFTKILRVQVRCAGALQLLGYRP 361
 Db 408 FALNMTSGSSSKP---FVVSARNATQAAANWARTALTFOQIKQVVEFCYQPMVILGYER 464
 QY 362 VYSADQOORLTLDLV 376
 Db 465 VNSPEVVKDLSTLL 479

RESULT 12
 Q9Y4C5 PRELIMINARY; PRT; 531 AA.
 ID Q9Y4C5
 AC Q9Y4C5
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE LONG FORM OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE
 DE (GLCNAC6ST).
 GN GN6ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RX MEDLINE=98391845; PubMed=9722682;
 RA Uchimura K., Muramatsu H., Kaname T., Ogawa H., Yamakawa T., Fan Q.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K.,
 RA Ozaki T., Nakagawara A., Kadomatsu K., Muramatsu T.,
 RT "Human N-acetylglucosamine-6-O-sulfotransferase involved in the
 RT biosynthesis of 6-sulfo sialyl Lewis X: Molecular cloning, chromosomal
 RT mapping, and expression in various organs and tumor cells.";
 RL J. Biochem. 124:670-678(1998).
 DR EMBL: AB014680; BAA34266.1; --
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferase.
 SQ SEQUENCE 531 AA; 58768 MW; D2EDB74E95B5162F CRC64;

Query Match
 Best Local Similarity 39.7%; Pred. No. 2.2e-53;
 Matches 149; Conservative 68; Mismatches 121; Indels 37; Gaps 12;

QY 34 SPAGGEDRVH-VLVLSWRSGSFLGQLFSQHPDVFYLMPEPAHWVWTLTGSGSAATLHMA 92
 Db 157 APEGVGDKRHWMYVFTTWRSSTGSGSFFGELFNQNPFFLYEPVHWVWQKLYPGDAVSLQGA 216
 QY 93 VRDLMSRIFLCMDVFDAYMPQ---SRNLS--AFFNWAISRALCSPACSAFFRGTTISKQ 147

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=PLACENTA;
RC	MEDLINE=20498786; PubMed=11042394;
RA	Sakaguchi H., Kitagawa H., Sugahara K.;
RT	"Functional expression and genomic structure of human N-
RT	acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-
RT	acetylglucosamine at the nonreducing end of an N-acetylglucosamine
RT	sequence.";
RL	Biochim. Biophys. Acta 1523:269-276(2000).
DR	EMBL; AB021125; BAB16887.1; -
DR	EMBL; AB021124; BAB16886.1; -
KW	Transferase.
SQ	SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;
Query Match 32.2%; Score 659.5; DB 4; Length 483;	
Best Local Similarity 39.1%; Pred. No. 1e-53;	
Matches 151; Conservative 68; Mismatches 122; Indels 45; Gaps 12;	
QY	31 GPSSPAG-----GDRVHVLVLSWRSGSFLGQFSQHPDVFYLMPEPAHVWTL 81
Db	98 GVAAPGNGTRGTGGVGGDKRQLVYVFTWRSQSGSFFGELEFNQNPVEFFLYEPVHWQKL 157
QY	82 SQGSAAATLHMAVRDLMRSIFPLCDMDYFDAYMPQ---SRNLS--AFFNWTASALCSPAC 136
Db	158 YPGDAVSLQGAARDMLSALYRCDLSVPQLYSPAGSGGRNLTTLGIFGAANVKVCSPLC 217
QY	137 SAPRGTISKQD--VCKTCTROPFSLAREACRSYSHVLKEVRFNQLVPLSDPAL 194
Db	218 PAYRKEVGVLDVRVCKK--CPQRLARFECECKYRVLVGVNVDVAVLAPLRDPAL 276
QY	195 NLRTVHLVRDPRVLR-----REAAAPILARD-----NGVLGTCKWY 234
Db	277 DLKVIHLVRDPRVLR-----REAAAPILARD-----NGVLGTCKWY 336
QY	235 --EADPH--LRLIRVCRSHVRIAEAAATLKPEPLRGYRLVRFEDLAREPILALYAF 291
Db	337 GGPADYHALGAMVEICNSMAKTQTA--LQPPDWLQGHYLVVRYEDLVGDPVKTLLRVYDF 395
QY	292 TGLTLPQLEAWIHNIHTHSG--IGKPIEAFTSSRNARNVSOAWRHLPFTKILRVQVVC 350
Db	396 VGLVSPDMEQFALNMISGSSSKP---FVVSARNATQAAAWRTALTFOQIKQVEEFC 452
QY	351 AGALQLLGYREYVSADQQRDLTLV 376
Db	453 YQPMVLYGVRSNPEEVKDLSTKLL 478
RESULT 10	
QY6F2	PRELIMINARY; PRY; 530 AA.
ID	QY6F2
AC	QY6F2;
DT	01-NOV-1999 (TrEMBLrel. 12, Created)
DT	01-NOV-1999 (TrEMBLrel. 12, last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, last annotation update)
DE	CARBOHYDRATE SULFOTRANSFERASE 2.
GN	CHST2.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RC	MEDLINE=99168906; PubMed=10049591;
RA	Li X., Tedder T.F.;
RT	"CHST1 and CHST2 sulfotransferases expressed by human vascular
RT	endothelial cells: cDNA cloning, expression, and chromosomal
RT	localization.";
RT	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RC	MEDLINE=99168906; PubMed=10049591;
RA	Li X., Tedder T.F.;
RT	"CHST1 and CHST2 sulfotransferases expressed by human vascular
RT	endothelial cells: cDNA cloning, expression, and chromosomal
RT	localization.";
RT	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=UMBILICAL VEIN ENDOTHELIAL CELLS;
RC	MEDLINE=99168906; PubMed=10049591;
RA	Li X., Tedder T.F.;
RT	"CHST1 and CHST2 sulfotransferases expressed by human vascular
RT	endothelial cells: cDNA cloning, expression, and chromosomal
RT	localization.";
RT	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=WHOLE EMBRYOS;
RC	MEDLINE=96380482; PubMed=9712885;
RA	Uchimura K., Muramatsu H., Kadomatsu K., Fan Q., Kurosawa N.,
RA	Mitsuka C., Kannagi R., Habuchi O., Muramatsu T.;
RT	"Molecular cloning and characterization of an N-acetylglucosamine-6-O-
RT	sulfotransferase.";
RL	J. Biol. Chem. 273:22577-22583(1998).
DR	EMBL; AB011452; BAA32138.1; -
DR	EMBL; AB011452; BAA32139.1; -
DR	EMBL; AB011451; BAA32137.1; -
DR	MGP; MGI:1891160; Chst2.
DR	InterPro; IPR000863; Sulfotransferase.
DR	Pfam; PF00685; Sulfotransferase; 1.
KW	Transferase.
SQ	SEQUENCE 530 AA; 57814 MW; A113ELP735C363EC CRC64;
Query Match 32.4%; Score 665; DB 11; Length 530;	
Best Local Similarity 40.4%; Pred. No. 3.4e-54;	
Matches 150; Conservative 64; Mismatches 121; Indels 36; Gaps 11;	
QY	37 GGRDVRVHVLVLSWRSGSFLGQFSQHPDVFYLMPEPAHVWTL 96
Db	160 GGDKRLVYVFTWRSQSGSFFGELEFNQNPVEFFLYEPVHWQKLPGDAVSLQGAARD 219
QY	97 MRSIFLDMVDYDAYMPQ---SRNLS--AFFNWTASALCSPACSAFPRGTISKQD--V 149
Db	220 LSALYRCDLSVFLYSPAGSGGRNLTTLGIFGAATNKVWCSSPLCPAYRKEVGLVDDRV 279
QY	150 CKTLCFOPFSLAREACRSYSHVLKEVRFNQLVPLSDPALNRLVHLVRDPAVL 209
Db	280 CKK--CPQRLARFECECKYRVLVGVNVDVAVLAPLRDPALDKVHLVRDPAVA 338
QY	210 RS-----REAAAPILARD-----NGVLGT--NGKWEADPH--LRLIREV 246
Db	339 SSRIRSHGLIRSLQVRSRDPRAHMPFLAAGHKGAKKEGMPADYHALGAMVEI 398
QY	247 CRSHVRIAEATLKPPFELRGYRLVRFEDLAREPILALYAF 306
Db	399 CNSMAKTQTA--LQPPDWLQGHYLVVRYEDLVGDPVKTLLRVYDFVGLVSPDMEQFALN 457
QY	307 ITHGSG--IGKPIEAFTSSRNARNVSOAWRHLPFTKILRVQVVCAGALQLLGYRVPVSA 365
Db	458 MTSQSGSSSKP---FVVSARNATQAAAWRTALTFOQIKQVEEFCYQPMVLYGVRSN 514
QY	366 DQQRDLTLV 376
Db	515 EEVKDLSTKLL 525
RESULT 9	
QY6Z5	PRELIMINARY; PRY; 483 AA.
ID	QY6Z5
AC	QY6Z5;
DT	01-MAR-2001 (TrEMBLrel. 16, Created)
DT	01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT	01-MAR-2001 (TrEMBLrel. 16, last annotation update)
DE	N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.
OS	Homo sapiens (Human).

```

QY 357 LGYRPVYSADQOQDLTDLVLPQDPDHFESWASPD 390
Db 357 LGYRHRVSRSEQORNLDDLL-----STWTVPE 383

RESULT 6
Q9RII1
ID Q9RII1 PRELIMINARY; PRT; 386 AA.
AC Q9RII1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE L-SELECTIN LIGAND SULFOTRANSFERASE.
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9361934; PubMed=10435581;
RA Hiraoka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";
RL Immunity 11:79-89(1999).
DR EMBL; AF109155; A045579.1; -.
DR MGD; MGI:1349479; Chst4.
KW Transferase; Lectin; Selectin.
SQ SEQUENCE 388 AA; 44636 MW; 6D5371AFB6884ABE CRC64;

Query Match 47.18; Score 965.5; DB 11; Length 388;
Best Local Similarity 55.9%; Pred. No. 1e-82;
Matches 194; Conservative 41; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVLLVSWRSGSFLGOLFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSRF 101
Db 42 VHVLLVSWRSGSFLGOLFQHPDVFYLMPEPAWHVWTTLSQGSAAATLHMAVRDLMSRF 101

QY 102 LCDMDVFAYM-POSRNLSAFNNWATSRALCPACAPRGITISKQDVCKTLCTRPQFS 160
Db 102 LCDMSVFDAYMNPGRKQSSLFQEQSALCSAPVCDPFPFAHEISSPKHKLICGQQPF 161

QY 161 LARACRSYSHVWLKEVFENLVLPDPAALNLRVHLVDRPRAVRSREAGPILA 220
Db 162 MVEKACRSHGVVLKEVFLSLQALYPLTDPDLNLRVHLVDRPRAVRSREAGPILA 221

QY 221 RDNGIVLTNGKWEADPHLRLE-----VCRSHVRIAEATLKPFPFLRGY 269
Db 222 VDSHVLG-----QHLETIKEEDQPYAMKTIKCSQDVIVKAIQTL--PEALQRY 270

QY 270 RLVPEDLAREPLAERIALYAFGTGLTLPQLEAWHTHSGSGTKPKTEAFHTSSRNARN 329
Db 271 LFLRYEDLVRAPLAQTLRLYKFGVGLDEPLQLQTVWVHVNTRGKMGQ--HAFETNARNALN 328

QY 330 VSOAQRHALPTKILRVCEVCAGALQLLGYRPVYSADQOQDLTDLV 376
Db 329 VSOAQRWSLPYKVSQLODAGEAMDLLGLYQVRSQEQGNLSDLL 375

RESULT 7
Q9WUE5
ID Q9WUE5 PRELIMINARY; PRT; 388 AA.
AC Q9WUE5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE N-ACETYLGUCOSAMINE 6-O-SULFOTRANSFERASE (CARBOHYDRATE (CHONDROITIN
DE 6/KERATAN) SULFOTRANSFERASE 4).
GN CHST4.
OS Mus musculus (Mouse).

```

```

QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRYQVCECAGALQLLYR 360
Db 300 EAWIHNTGSGPGARREAFKYSRNALNVQAWRHALPFAKIRRYQVCECAGALQLLYR 359
QY 361 PYSADQQRDLTDLVLRPGDPFWSAS 388
Db 360 PYSDEQQRNALDVLPRGLNGFTWAS 387

RESULT 4
Q90UP4 PRELIMINARY; PRT; 395 AA.
AC Q90UP4;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE.
GN CHST5 OR 1-GLCNAC-6-ST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=99423499; PubMed=10491328;
RX STRAIN=C57BL/6; TISSUE=INTESTINE;
RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-
RT Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
DR EMBL; AF176841; AAD56003.1; -.
DR EMBL; AF176840; AAD56002.1; -.
DR MGD; MGI:1931825; Chst5.
KW Transferase.
SQ SEQUENCE 395 AA; 44537 MW; 3DF71E43ED383BE CRC64;

Query Match 74.9%; Score 1536.5; DB 11; Length 395;
Best Local Similarity 76.0%; Pred. No. 1.5e-136;
Matches 295; Conservative 30; Mismatches 62; Indels 1; Gaps 1;

QY 1 MMLPFSSTVIVLLLAQTCLLLFTISRPSPAGGEDRVHVLVSSWRSGSFLGQL 60
Db 1 MRLPFSSTVIVLLLAQTCLLLFTISRPSPAGGEDRVHVLVSSWRSGSFLGQL 59
QY 61 FSQHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLMSRSTFLCDMDVFDAYMPOSRLNSA 120
Db 60 FSQHPDVFYLMPEAWHVTTLTSGSAPALHMAVRDLMSRSTFLCDMDVFDAYLPWRNNSD 119
QY 121 FENWATSRALCSPACSAFPGTISKQDVCKTLCTRPFSLAEEACRSYSHVVLKEVRFF 180
Db 120 LFQWAVSRALCSPVCEAFARGNISSECKPLCATRPFGLAQEACSSYSHVVLKEVRFF 179
QY 181 NLQVLPILLSDPALNLRIVHVRDPAVLRSREAAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLPILLSDPALNLRIVHVRDPAVLRSREAAAGPILARDNGIVLGTNGTWEADPHL 239
QY 241 RLIREVCRSHVRIAEAAATKPPFELRGYRLVRFEDLAREPLAEIRALYAFGLTGLTPQL 300
Db 240 RYVNEVCRSHVRIAEAAALHKPPFELQDRYRLVRYEDLARDPLTVIRELYAFGLTGLTPQL 299
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRYQVCECAGALQLLYR 360
Db 300 QWIIHNTGSGPGARREAFKITSRLNVQAWRHALPFAKIRRYQVCECAGALQLLYR 359
QY 361 PYSADQQRDLTDLVLRPGDPFWSAS 388
Db 360 SVHSELEQRDLSDLLPRGDMFSKWSAS 387

RESULT 5
Q9Y5R3 PRELIMINARY; PRT; 386 AA.
ID Q9Y5R3

```

```

AC Q9Y5R3;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE (L-SELECTIN LIGAND
DE SULFOTRANSFERASE GST-3).
GN GST3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal Localization and Genomic Organization for the
RT Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-
RT Sulfotransferase Gene Family.";
RL Glycobiology 0:0-0(2001).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Elies L.G.,
RA Rabuka D., Hindsgaul O., Marth J.D., Lowe J.B., Fukuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core1 extension betal,3-N-acetylglucosaminyltransferase.";
RL Cell 105:957-969(2001).
DR EMBL; AF131235; AAD33015.1; -.
DR EMBL; AF280088; AAG48246.1; -.
DR EMBL; AF149783; AAK48417.1; -.
DR InterPro; IPR003015; HLH_Myc.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KW Transferase; Lectin; Selectin.
SQ SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

```

Query Match 49.1%; Score 1008; DB 4; Length 386;
 Best Local Similarity 52.3%; Pred. No. 9.9e-87;
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

```

QY 14 LLLAQTCLLLFTISRP-----GPSPPAGGEDRVHVLVSSWRSGSFLGQ 59
Db 1 MLLPKMKLLFLVLSQMAILALFFHMYSHNTSSLSMKAAQPERMHVILVSSWRSGSFLGQ 60
QY 60 LFSQHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLMSRSTFLCDMDVFDAYM-POSRLN 118
Db 61 LFGQHPDVFYLMPEAWHVTTLTSGSAAATLHMAVRDLMSRSTFLCDMDVFDAYMPPGPRQ 120
QY 119 SAFENWATSRALCSPACSAFPGTISKQDVCKTLCTRPFSLAEEACRSYSHVVLKEVR 178
Db 121 SSFLFOWNSRALCSPACSAFPGTISKQDVCKTLCTRPFSLAEEACRSYSHVVLKEVR 180
QY 179 FNNLQVLPILLSDPALNLRIVHVRDPAVLRSREAAAGPILARDNGIVLGTN-GKWVEAD 237
Db 181 FNNLQVLPILLSDPALNLRIVHVRDPAVLRSREAAAGPILARDNGIVLGTN-GKWVEAD 240
QY 238 PHLRLIREVCRSHVRIAEAAATKPPFELRGYRLVRFEDLAREPLAEIRALYAFGLTGL 296
Db 241 QPYVVMQVCSQLEIYK--TIQSLKALQRYLLVRYEDLARDPAVACTSRMYFEVGLF 298
QY 297 TPQLEAWIHNTGSGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRYQVCECAGALQL 356
Db 299 LPHLQIWWHNTGSGIGKPIEAFHTSSRNARNVQAWRHALPFTKILRYQVCECAGALQL 356

```



```

QY 121 FFWATSRALCSPACSAFPGTISKQDVCKTLCITROPFSLAREACRSYSHVVLKEVRF 180
Db 121 FFWATSRALCSPACSAFPGTISKQDVCKTLCITROPFSLAREACRSYSHVVLKEVRF 180
QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Db 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
QY 241 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 300
Db 241 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 300
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLGYR 360
Db 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLGYR 360
QY 361 PYSADQQRDLTLDVLPRGPDHFSWSPD 390
Db 361 PYSADQQRDLTLDVLPRGPDHFSWSPD 390
RESULT 2
Q9GZS9 PRELIMINARY; PRT; 411 AA.
AC Q9GZS9;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE INTESTINAL GLCNAC-6-SULFOTRANSFERASE (INTESTINAL N-
DE ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).
GN CHST5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A.,
RA Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF246718; AAG28023.1; -;
DR EMBL; AF219991; AAG26326.1; -;
KW Transferase.
SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match 100.0%; Score 2051; DB 4; Length 411;
Best Local Similarity 100.0%; Pred. No. 4.5e-185;
Matches 390; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MWLPFSKTVTVLLAQTCLLFIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLGQL 60
Db 22 MWLPFSKTVTVLLAQTCLLFIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLGQL 81
QY 61 FSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVFDAYMQSRNLSA 120
Db 82 FSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVFDAYMQSRNLSA 141
QY 121 FFWATSRALCSPACSAFPGTISKQDVCKTLCITROPFSLAREACRSYSHVVLKEVRF 180
Db 142 FFWATSRALCSPACSAFPGTISKQDVCKTLCITROPFSLAREACRSYSHVVLKEVRF 201
QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Db 202 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 261
QY 241 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 300
Db 241 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 300

```

```

Db 262 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 321
QY 301 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLGYR 360
Db 322 EAWIHNTGSGIGKPIEAFHTSSRNARNVSOAWRHALPFTKILRVQEVCAQALQLGYR 381
QY 361 PYSADQQRDLTLDVLPRGPDHFSWSPD 390
Db 382 PYSADQQRDLTLDVLPRGPDHFSWSPD 411
RESULT 3
Q9GZX3 PRELIMINARY; PRT; 395 AA.
AC Q9GZX3;
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE GST-4BETA (CORNEAL N-
DE ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE).
GN GST4BETA OR CHST6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal Localization and Genomic Organization for the
RT Galactose/N-Acetylgalactosamine/N-Acetylglucosamine 6-O-
RT Sulfoltransferase Gene Family.";
RL Glycobiology 0:0-0(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20472330; PubMed=11017086;
RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,
RA Nakamura T., Dota A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;
RT "Macular corneal dystrophy type I and type II are caused by distinct
RT mutations in a new sulphotransferase gene.";
RL Nat. Genet. 26:237-241(2000).
DR EMBL; AF280086; AAG48244.1; -;
DR EMBL; AF219990; AAG26325.1; -;
DR EMBL; AF219991; AAG26327.1; -;
KW Transferase.
SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48F67 CRC64;

```

```

Query Match 84.3%; Score 1729.5; DB 4; Length 395;
Best Local Similarity 85.8%; Pred. No. 9.1e-155;
Matches 333; Conservative 17; Mismatches 37; Indels 1; Gaps 1;

```

```

QY 1 MWLPFSKTVTVLLAQTCLLFIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLGQL 60
Db 1 MWLPFSKTVTVLLAQTCLLFIISRPGPSPPAGGEDRVHVLVLSWSRSGSFLGQL 59
QY 61 FSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVFDAYMQSRNLSA 120
Db 60 FSQHPDVLYLMEPAWHVWTTLSQGSAAATLHMAVRDLMSIFLCMDVFDAYMQSRNLSA 119
QY 121 FFWATSRALCSPACSAFPGTISKQDVCKTLCITROPFSLAREACRSYSHVVLKEVRF 180
Db 120 FFWATSRALCSPACSAFPGTISKQDVCKTLCITROPFSLAREACRSYSHVVLKEVRF 179
QY 181 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 240
Db 180 NLQVLYPLSDPALNLRIVHLVRDPRAVLRSRRAAGPILARDNGIVLGTNGKWEADPHL 239
QY 241 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 300
Db 240 RLIREVCSHVRIRAEAAATLKPPPLRGYRLVRFEDLAREPLAEIRALYAFTGLITPOL 299

```

